TECHNICKÁ UNIVERZITA V LIBERCI
Fakulta mechatroniky, informatiky a mezioborových studií

Vliv nahrávacího řetězce na identifikaci hudební nahrávky

Bakalářská práce

Studijní program: B2646 – Informační technologie
Studijní obor: 1802R007 – Informační technologie

Autor práce: **David Václavek**
Vedoucí práce: Ing. Marek Boháč

Liberec 2014

TECHNICAL UNIVERSITY OF LIBEREC
Faculty of Mechatronics, Informatics
and Interdisciplinary Studies

The impact of the recording chain on the identification of a musical piece

Bachelor thesis

Study programme: B2646 – Informational Technologies
Study branch: 1802R007 – Informational Technologies

Author: **David Václavek**
Supervisor: Ing. Marek Boháč

Liberec 2014

Tento list nahraďte
originálem zadání.

Prohlášení

Byl jsem seznámen s tím, že na mou bakalářskou práci se plně vztahuje zákon č. 121/2000 Sb., o právu autorském, zejména § 60 – školní dílo.

Beru na vědomí, že Technická univerzita v Liberci (TUL) nezasahuje do mých autorských práv užitím mé bakalářské práce pro vnitřní potřebu TUL.

Užiji-li bakalářskou práci nebo poskytnu-li licenci k jejímu využití, jsem si vědom povinnosti informovat o této skutečnosti TUL; v tomto případě má TUL právo ode mne požadovat úhradu nákladů, které vynaložila na vytvoření díla, až do jejich skutečné výše.

Bakalářskou práci jsem vypracoval samostatně s použitím uvedené literatury a na základě konzultací s vedoucím mé bakalářské práce a konzultantem.

Současně čestně prohlašuji, že tištěná verze práce se shoduje s elektronickou verzí, vloženou do IS STAG.

Datum:

Podpis:

Abstrakt

Tato bakalářská práce se zabývá návrhem systému pro rozpoznávání hudebních nahrávek z databáze a vlivem nahrávacího řetězce na úspěšnost takového systému. Databáze hudebních děl je reprezentována skupinou vhodně zvolených příznaků. Těmi jsou tempo nahrávky a dominantní frekvence jednotlivých dob skladby. V úvahu jsou brány reálné nahrávací podmínky, jako akustika místností a přenosové charakteristiky zařízení, používaných pro reprodukci skladby a nahrávání. Úspěšnost systému a vliv nahrávacího řetězce jsou experimentálně vyhodnoceny.

Klíčová slova: systém rozpoznávání písní, příznaky písně, vliv nahrávacího prostředí, přenosové charakteristiky

Abstract

This Bachelor's thesis deals with a proposal of a system for a musical piece identification and with the impact of the recording chain. The musical piece database is represented by sets of chosen features, such as tempo and dominant frequencies of each musical period. Real recording conditions are taken into account - the room impulse response and transfer characteristics of the impoyed devices. The accuracy of the system and the impact of the recording chain are experimentally evaluated.

Keywords: music recognition system, song signatures, recording environment influence, transfer characteritics

Poděkování

V první řadě bych velmi rád poděkoval Ing. Marku Boháčovi za cenné rady, věcné připomínky a vstřícnost při konzultacích a vypracování bakalářské práce. Můj dík dále patří Ing. Jiřímu Málkovi, Ph.D. a Ing. Michalu Rottovi za ochotu a pomoc při řešení problémů. Chtěl bych poděkovat mé rodině, která mě ve stresových chvílích dokázala podržet a nabudit do další práce. Dík patří i přítelkyni, především za její trpělivost a pevné nervy. V poslední řadě bych chtěl poděkovat všem přátelům, kteří se psychickou podporou také zasloužili o úspěšné dokončení této práce.

Obsah

Seznam obrázků

Seznam tabulek

Seznam zkratek

FIR Finite Impulse Response (konečná impuzní odezva, např.: filtru)

WAVE, WAV Waveform Audio File Format (typ hudebního formátu)

BPM Beat Per Minute (počet úderů bubeníka za minutu)

PC Personal Computer (osobní počítač)

Mac Macintosh (rodina osobních počítačů od Apple, Inc.)

Hz Hertz (základní jednotka kmitočtu)

LTI Linear Time-Invariant (lineární a časově nezávislý, např.: systém)

IIR Infinite Impulse Response (nekonečná impuzní odezva, např.: filtru)

STFT Short-Time Fourier Transform (Krátkodobá Fourierova transformace)

IBM International Business Machines Corporation (společnost v oboru informačních technologií)

RIFF Resource Interchange File Format (multimediální kontejner pro ukládání multimediálních zvukových a obrazových souborů)

GNU GNU's Not Unix! (licence svobodného softwaru)

dB decibel (jednotka popisující intenzitu zvuku)

DVD externí médium pro ukládání dat

GB gigabyte (jednotka množství informace v informatice)

FTT Fast Fourier Transform (efektivní algoritmus pro spočtení diskrétní Fourierovy transformace)

IFFT Inverse Fast Fourier Transform (algoritmus pro spočtení zpětné diskrétní Fourierovy transformace)

HP Hewlett-Packard (společnost zabývající se informačními technologiemi)

MS Excel tabulkový procesor od firmy Microsoft

PDF Portable Document Format (platformě nezávislý formát dokumentu)

LaTeX typografický systém, soubor maker pro sázecí program TeX

Úvod

Úkolem této bakalářské práce je navrhnout systém pro rozpoznávání písní z databáze. V potaz jsou brány vlivy reálného nahrávacího prostředí, jako akustika místnosti a přenosové vlastnosti testovaných zařízení. Systém bude pracovat na principu rozpoznání tempa, na základě kterého se poté určí dominantní frekvence dob písně. Úspěšnost navrženého systému bude poté experimentálně vyhodnocena.

Motivace

Toto téma jsem si vybral z prostého důvodu. Tyto systémy jsou ve velkém množství komerční, a proto jsem chtěl nejen sobě, ale i ostatním osvětlit tuto problematiku a nástin možného návrhu samotného řešení. Systém se skládá z podsystémů, které mohou samy o sobě usnadnit hudebníkům práci a lépe tak oslovit ucho posluchače.

Přehled existujících technologií

Existuje nemalé množství technologií[1] [2], ať už na mobilní nebo i desktopové platformy, které byly vyvinuty za účelem rozpoznávání písní. Velké procento však zastupují programy komerční. To ve většině případů znamená, že počet rozpoznání je na určitý časový úsek omezen. Za příplatek si lze pořídit jejich prémiovou verzi, ve které se zpřístupní všechny dostupné funkce.

Kód programu není samozřejmě volně dostupný, a tak lze přesnější princip funkčnosti programu odhadovat pouze z obecného popisu, zveřejněného výrobcem. Softwary pracují na různých principech. Spoléhají na kombinace rozpoznání použitých hudebních nástrojů, tempa, basové linky, akordů, žánru nebo například konkrétních tónů skladby.

Následuje přehled nejznámějších programů pro rozpoznávání písní. Existují i další, které však ve světě nemají takové jméno a jejich rozpoznávací schopnosti nejsou tak dobré.

[1] Čerpaný přehled (1): mashable.com/2010/03/30/identify-song-apps

[2] Čerpaný přehled (2): evolver.fm/2012/10/10/top-5-apps-for-identifying-songs

Midomi

Midomi je webová aplikace, která dokáže rozpoznat tóny nahrávky. Je tedy vhodná i pro případy, kdy Vám nějaká melodie uvízne v hlavě. Je však třeba, aby se člověk vyskytoval v prostředí bez rušivých elementů.

Nahrávání trvá přibližně 10 sekund. Po vyzkoušení broukání pěti písní různých žánrů (hip-hop, pop, rock, ska, soul) byla úspěšnost rozpoznání 40%.

Tunatic

Tunatic je volně dostupná klient-server aplikace pro PC a Mac. Nahrávání skladby probíhá jednoduše, tedy po stisku nahrávacího tlačítka. Opět, je třeba, aby prostředí, ve kterém se nahrává, bylo klidné.

I když program vrátil výsledky během pěti sekund, pouze jedna z pěti vyzkoušených skladeb byla rozpoznána správně. Úspěšnost rozpoznání byla tedy 20%.

Shazam

Shazam je komerční služba dostupná pro mobilní platformy i jako webová aplikace. Volně dostupná varianta Vám dovolí rozpoznat pět skladeb za měsíc. Z vlastní zkušenosti lze říct, že program slibně pracuje i v prostředí, kde se vyskytuje nežádoucí šum.

Nahrávání probíhá po dobu 10 sekund. Poté je nahrávka odeslána na server k vyhodnocení. Pro 5 skladeb program pracoval s úspěšností 100%.

SoundHound

Další komerční mobilní aplikace, která umí rozpoznat píseň broukáním nebo zazpíváním několika řádků textu.

Nahrávka má opět kolem 10 sekund a úspěšnost při stejném testu 60%.

Spotsearch

Spotsearch je mobilní aplikace, která rozpoznává písně pomocí textu. Funguje tak, že na základě textu určí neznámou píseň. Text je nutné ručně napsat.

Z pěti testovaných písní byly rozpoznány 4, tedy 80%.

Využití technologií, aplikace pro hudebníky

Vytvářený systém se zabývá problémem rozpoznávání písní. Kromě samotného zjištění názvu písně, což je stěžejním bodem, lze poté vhodným rozšířením systému dohledat a zobrazit text, notový zápis nebo například akordy písně. Je zřejmé, že i všechna tato data se dají z písně rozpoznat. Analýza těchto dat a návrh systémů je však komplikovaný.

Systém se dělí na části, které jsou samostatně funkčními bloky. Každý tento blok analyzuje signál a zjišťuje o něm určité informace, které by mohly řešit problémy sužující hudebníky. Nyní budou stručně popsány aplikace, které by mohly jednotlivé body práce řešit.

Analýza hry hudebníka

První variantou, kterou by mohl systém řešit, je analýza hry hudebníka. To znamená, pokud hudebník hraje svou část skladby, lze jeho hru analyzovat a vytvořit posloupnost not, které zahrál. Poté lze tóny porovnat s předlohou (notovým zápisem), podle kterého hrál. Lze tak zjistit, kde udělal chybu a příště se jí vyvarovat.

Analýza hry bubeníka

Tato aplikace by mohla pomoci v případech, když by bubeník nedokázal dodržet tempo písně. Aplikace by mohla bubeníka upozornit, ať už vizuálně na obrazovce nebo zvukově do sluchátek na to, aby dle potřeb upravil rychlost hry.

Metronom

Metronom v podstatě vychází z předchozího bodu. Je-li nutné, aby hudebník dodržel tempo, a nemá k dispozici bicí linku, která by ho vedla, je metronom dobrou volbou.

Ladění hudebních nástrojů

Klasickým případem aplikace je ladička hudebních nástrojů. Možností by mohlo být zvolení nástroje a typu ladění. Po zahrání noty (například chvění struny kytary či klavíru) by systém vyhodnotil, zda frekvence tónu sedí, případně, jak strunu nástroje upravit.

Další možností je počítání úderů za minutu, které se hodí například do soutěží bubeníků.

Automatické otáčení stránek partitury

Zajímavým využitím systému by mohlo být automatické otáčení stránek partitury. Problémem při ručním otáčení je, že hudebník může zazmatkovat a notový zápis shodit na zem. Pokud zrovna musí hrát, je zapotřebí další osoby, která mu s otáčením pomůže. Tyto nepříjemnosti by mohla řešit aplikace, která by detekovala zahrané noty a na konci stránky s notovým zápisem by otočila na další. Zápis by byl s největší pravděpodobností zobrazen na elektronickém zařízení s displejem.

Tento problém je již částečně vyřešen, avšak aplikace spoléhá na známé, pevně zadané tempo, čili funguje jako časovač.

1 Teoretické základy

1.1 Beat synchronous systémy

Beat synchronous systémy jsou systémy, které slouží ke klasifikaci zvukového signálu. Na rozdíl od technologií rozpoznávání řeči, které pracují s pevným oknem, mají tyto systémy jiný způsob segmentace. Jelikož slouží především ke klasifikaci hudebních signálů, dělí je na takty či případně doby. Hudební nahrávku tak separují do bloků, které jsou mezi sebou následně porovnávány a analyzovány. To přináší úsporu výpočetního výkonu, časovou úsporu a lepší výsledky [1].

1.2 Teorie zpracování signálů

1.2.1 Diskrétní signál, obdélníkový signál

Diskrétní signál (viz obrázek 1.1) je indexovaná posloupnost reálných nebo komplexních čísel. To znamená, že hodnota signálu je definována pouze v izolovaných okamžicích [2]. Také platí, že tento signál nabývá v čase pouze jedné z konkrétních hodnot. V praxi se tento signál nazývá také digitální signál.

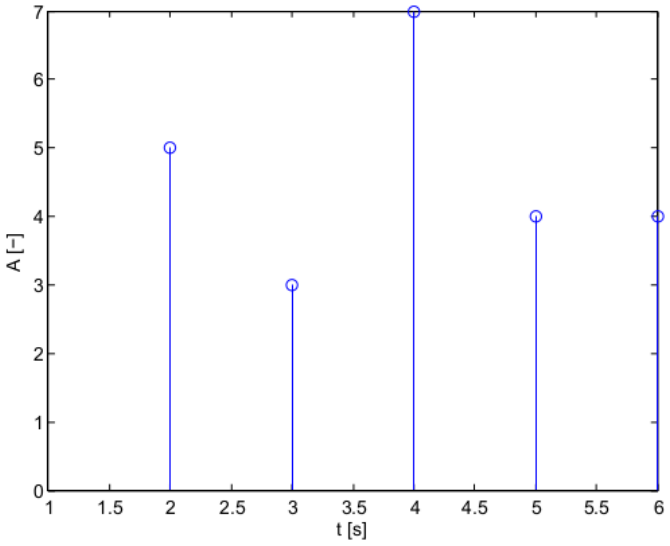
V mé práci je kromě klasických diskrétních signálů, jejichž analýzou a úpravou se zabývám, použita také podmnožina tohoto signálu, nazvaná obdélníkový signál [5]. Obdélníkový signál používám při výběru správného tempa hudební nahrávky a časovou detekci počátků jednotlivých dob.

Ideální obdélníkový signál nabývá v čase pouze dvou konstantních hodnot. Jsou to úrovně nula a jedna. V reálném světě signálů není díky parametrům digitálně analogových převodníků možné dokonalý obdélníkový signál vytvořit. Skládá se z nekonečného počtu sinusových vln, kde následující vlna obsahuje druhou vyšší harmonickou vlny předchozí. Parametrem tohoto signálu je pojem zvaný střída.

Střída signálu je poměr mezi délkou trvání hodnot obou úrovní. To znamená, že signál se střídou 1:4, má čtyřikrát kratší horní úroveň signálu, než úroveň dolní.

1.2.2 Vzorkovací frekvence, Shannonův teorém

Vzorkovací frekvence definuje počet vzorků za časový úsek jedné sekundy, načítaných ze spojitého analogového signálu při jeho přeměně na signál digitální. Jednotkou vzorkovacího kmitočtu je Hertz (značení Hz). Je-li tedy vzorkovací frekvence diskrétního signálu 16 kHz, je za jednu sekundu nashromážděno 16 000 vzorků.



Obrázek 1.1: Grafická reprezentace diskrétního signálu

Vzorkovací teorém (viz vztah 1.1), jinak také Shannonův či Nyquistův teorém uvádí, že dokonalou rekonstrukci signálu lze provést pouze tehdy, je-li vzorkovací frekvence alespoň dvakrát vyšší, než nejvyšší frekvence vzorkovaného signálu. Pokud by tento teorém nebyl dodržen, došlo by k takzvanému aliasingu, který způsobuje přeložení signálu takovým způsobem, kdy je frekvence původního signálu přeložena na jinou.

$$F_s \geq 2f_{max} \tag{1.1}$$

Kde:
F_s je vzorkovací frekvence [Hz],
f_{max} je nejvyšší frekvence vzorkovaného signálu [Hz].

1.2.3 Základní operace se signály v praxi

Pokud zvukový signál vynásobím libovolným číslem z oboru reálných čísel, změní svou úroveň hlasitosti (viz vztah 1.2). Je-li číslo, kterým signál násobím, větší, než 1, hlasitost se zvýší, v opačném případě naopak.

$$x_B = Ax_A \tag{1.2}$$

Kde:
x_B je signál s modifikovanou amplitudou [-],
A je koeficient pro madifikaci amplitudy [-],
x_A je původní signál [-].

Změnu hlasitosti, respektive její normování, používám při vytváření testovací databáze, kdy je po dokončení metody Overlap-Add třeba zamezit přebuzení amplitudy signálu.

Sečtení kanálů signálu prakticky způsobí, že se dvoukanálový (stereo) signál stane jednokanálovým (mono) signálem (viz vztah 1.3), přičemž kromě prostorového efektu budou všechny ostatní informace zachovány.

$$sig[n] = x[n, 1] + x[n, 2] \tag{1.3}$$

Kde:
$sig[n]$ je hodnota jednokanálového signálu v čase n [-],
$x[n, a]$ je hodnota signálu kanálu a v čase n, který chci sloučit [-].

Tuto vlastnost používám pro zjednodušení analýzy nahrávek, jelikož při reálném nasazení tohoto druhu systému se nepoužívá dvou mikrofonů pro nahrávání sterea. Další využití jsem nalezl při vytváření testovací databáze.

1.2.4 Energie signálu, okenní funkce

Energie (viz vztah 1.4) je vedle výkonu, střední hodnoty a rozptylu signálu hlavní charakteristickou veličinou signálů. Energie vzorku digitálního signálu je dána jako kvadrát vzorku signálu.

Pokud by byla energie vyjadřována z každého vzorku signálu, došlo by ke znatelnému nárůstu výpočetních nároků na systém. Proto je při výpočtech energie běžné používat tzv. okenní funkci. Ta zajistí, že je energie vypočítána na určitém počtu vzorků (určitém časovém úseku).

$$E[n] = \sum_{n=0}^{N} (x[n])^2 \tag{1.4}$$

Kde:
E je energie vzorku signálu [-],
$x[n]$ je vzorek signálu v čase n [-],
N je konečný počet vzorků [-].

V praxi lze správnou analýzou energie signálu hudební nahrávky například zjistit, kde byly zahrány hlasité úseky (basa, úder bubeníka). Právě tuto možnost využívám v mé práci k tomu, abych byl schopen za pomocí dalších kroků odlišit jednotlivé doby a určit správné tempo nahrávky.

1.2.5 Konvoluce signálu

Konvoluce signálu (viz vztah 1.5) je matematická operace mezi dvěma signály. Jejími základními vlastnostmi jsou komutativnost, asociativnost a distributivnost. Délka výsledné konvoluční funkce je rovna součtu délek obou signálů vstupujících do konvoluční funkce mínus jedna.

$$x(n) * h(n) = \sum_{k=0}^{N} x(k)h(n-k) \tag{1.5}$$

Kde:
$x(n) * h(n)$ je konvoluce signálu a impulzní odezvy [-],
$x(k)$ je hodnota vzorku signálu v čase k [-],
$h(n-k)$ je hodnota vzorku impulzní odezvy v čase $n-k$ [-].

Konvoluci energie hudební nahrávky a obdélníkového signálu o určité periodě používám k určení správného tempa a detekci začátků dob písně. Dále je využita při vytváření testovací databáze, kde simuluji echo místnosti (metoda Overlap-Add).

1.2.6 Korelace a autokorelace

Korelace (viz vztah 1.6) je další významnou operací v oblasti zpracování signálů. Určuje vzájemnou podobnost dvou signálů [6]. Pokud provádím korelaci na jednom signálu, nazývá se tato operace autokorelace. Autokorelací lze například zjistit, zda se určité úseky signálu periodicky opakují. Vztahy pro výpočet jsou skoro identické, v autokorelaci je pouze nahrazen druhý signál signálem prvním [7].

$$R_{fg}[n] = \sum_{k} f^*[k]g[n+k] \tag{1.6}$$

Kde:
$R_{fg}[n]$ je korelační funkce [-],
$f*[k]$ je hodnota vzorku signálu f v čase k (komplexní sdružení) [-],
$g[n+k]$ je hodnota vzorku signálu g v čase $n+k$ [-].

Autokorelační funkci využívám při detekci tempa analyzované hudební nahrávky. Základní melodie nahrávky se totiž prakticky opakuje po každé době, která je charakterizována úderem bubnu či zvukem basy.

1.2.7 Jednotkový impulz, impulzní odezva systému

Jednotkový impulz lze popsat jako funkci, jejíž hodnota v čase nula je rovna jedné. V každém jiném čase je hodnota nulová. Při reálném měření impulzních odezev místností bývá nahrazován podobným krátkým a rázným signálem, například zabouchnutím dveří, tlesknutím, výstřelem z pistole. Za možnou variantu tohoto impulzu lze také považovat úder bubnu.

Přivedu-li na systém, za předpokladu nulových počátečních podmínek, jednotkový impulz, bude za impulzní odezvu systému považován výstupní signál. Určím-li tedy systémem školní učebnu a bouchnutí dveří jednotkovým impulzem, bude výsledná nahrávka považována za impulzní odezvu systému. Impulzní odezva systému je nejdůležitější charakteristikou LTI systémů [6]. LTI systém je takový systém, který v čase nemění své chování.

Impulzní odezvu systému používám při simulaci reálných nahrávacích podmínek, kde na signál aplikuji echo místnosti.

1.2.8 FIR filtry, sčítání filtrů

FIR filtr je typ filtru, který má konečnou impulzní odezvu. To znamená, že odezva filtru na konečný signál bude konečná. Filtr je popsán vztahem 1.7. Tento filtr je stabilní, jednoduchý k návrhu, nerekurzivní, ale jeho průběh je vzdálen ideálním filtrům. Opakem tohoto filtru je IIR filtr, tedy filtr s nekonečnou impulzní odezvou.

$$y[n] = h[0]x[n] + h[1]x[n-1] + ... + h[N-1]x[n-N-1] \tag{1.7}$$

Kde:
$y[n]$ je výstupní signál [-],
$h[n]$ je impulzní odezva filtru [-],
$x[n]$ vstupní signál [-],
N je řád filtru [-].

FIR filtraci využívám pro simulaci reálných podmínek, respektive kdy je třeba aplikovat přenosové vlastnosti každé kombinace nahrávacích zařízení na celou databázi hudebních děl. Další využití jsem nalezl při analýze tempa nahrávky, kdy je třeba obecně filtrovat skladby tak, aby byly přeneseny frekvence, které jsou společné pro všechna testovaná zařízení.

FIR filtry mají další vlastnost, kterou jsem využil vytváření testovací databáze. Lze provést jejich konvoluci, popřípadě sčítat jejich impulzní odezvy a tím docílit aplikování obou filtrů na nahrávku zároveň. Tato vlastnost platí jak pro paralelní zapojení, tak pro sériovou kombinaci.

1.2.9 Diskrétní Fourierova transformace, krátkodobá Fourierova transformace

Diskrétní Fourierova transformace je vyjádření časově závislého signálu pomocí harmonických signálů. Slouží pro převod diskrétních signálů z časové oblasti do oblasti frekvenční [3].

Rychlá Fourierova transformace je efektivní algoritmus pro spočtení diskrétní Fourierovy transformace (viz vztah 1.8). Transformace je důležitá zejména v oblastech digitálního zpracování signálu, ale nalezla své uplatnění při řešení parciálních diferenciálních rovnic [4].

$$X[k] = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{\frac{-j2\pi nk}{N}} \tag{1.8}$$

Kde:
$X[k]$ je výstup N hodnot komplexních koeficientů spektra [-],
N je počet hodnot číslicového spektra [-],

$x[n]$ je vstupní číslicový signál [-].

Jako krátkodobá Fourierova transformace je označována Fourierova transformace aplikovaná na analyzovanou funkci postupně po krátkých úsecích. Ty jsou určeny pomocí symetrického okna. Řeší se tím problém souběžného určení času i frekvence, na kterých je rozmístěna energie signálu.

STFT používám při spektrální analýze nahrávek, konktrétně při zjišťování dominantních frekvencí každé doby, které na základě dalšího postupu porovnávám s jinou nahrávkou.

1.2.10 Metoda Overlap-Add

Při zpracování signálů je třeba často provádět vzájemnou diskrétní konvoluci signálu a konečnou impulzní odezvou nějakého systému. Problémem však je, že čím je signál delší, tím je výpočetní náročnost a doba jeho trvání mnohonásobně vyšší.

Právě k potlačení těchto nevýhod slouží metoda Overlap-Add. Principem metody je rozdělit dlouhý signál na menší části, které se vzájemně nepřekrývají. Překryv je závislý na délce impulzní odezvy FIR systému. Odezva je převedena do frekvenční oblasti. Části signálu jsou také postupně převáděny do frekvenční oblasti, kde je provedena vzájemná konvoluce obou signálu. Následně je aplikován zpětný převod do časové oblasti.

Tuto metodu používám při vytváření testovací databáze, kdy simuluji impulzní odezvu místností, čili reálné nahrávací podmínky. Podrobný popis průběhu metody bude vysvětlen na odpovídajícím místě při popisu vlastní práce.

1.2.11 Formát souboru WAVE

WAVE (nebo také WAV) je zvukový formát, který byl vytvořen firmami IBM a Microsoft. Slouží k ukládání zvuku na počítači a je to varianta kontejneru RIFF. Formát RIFF umožňuje ukládat do souboru WAVE zvuk v různých variantách.

Jelikož je to bezeztrátový formát, je jeho zpracování snadné a výpočetně nenáročné. Proto se používá jako pracovní formát při zpracování zvuku. Velikost WAV souboru je omezena na čtyři gigabajty, což odpovídá šest a půl hodinám záznamu v CD kvalitě.

Pro jeho jednoduchost, snadné operace a plnou podporu v Matlabu jsem ho vybral pro uložení vzorové databáze a vývoj testovací databáze mé bakalářské práce.

2 Návrh systému pro identifikaci nahrávky

2.1 Celkový pohled na průběh rozpoznávání

Celý proces rozpoznávání neznámé hudební nahrávky z databáze příznaků skladeb známých nahrávek se skládá z několika netriviálních dílčích kroků, které popisuje diagram 2.1. Každý tento krok, bude podrobně popsán v dalších částech práce a bude mu věnována celá samostatná podkapitola.

Obecně je třeba, aby byla nejprve vybrána vzorová databáze písní, pokrývající pokud možno co největší okruh hudebního spektra. Tedy různé hudební žánry, rychlá s pomalou hudbou, instrumentální hudba a zpívaná hudba, zpěváci obou pohlaví, různé jazyky, atd.

Dalším krokem bylo třeba pokrýt nejčastější reprodukční a nahrávací zařízení z běžného života, tedy repro soustavy různých sestav, typů, mobilní telefony, notebooky, reproduktory projektorů, případně mikrofony sluchátek. Každé takové zařízení má své specifické přenosové vlastnosti, které musíme brát v potaz, a proto jejich vzájemnou kombinací můžeme docílit zobecnění úlohy a otestovat tak funkčnost navrhnutého řešení napříč zařízeními.

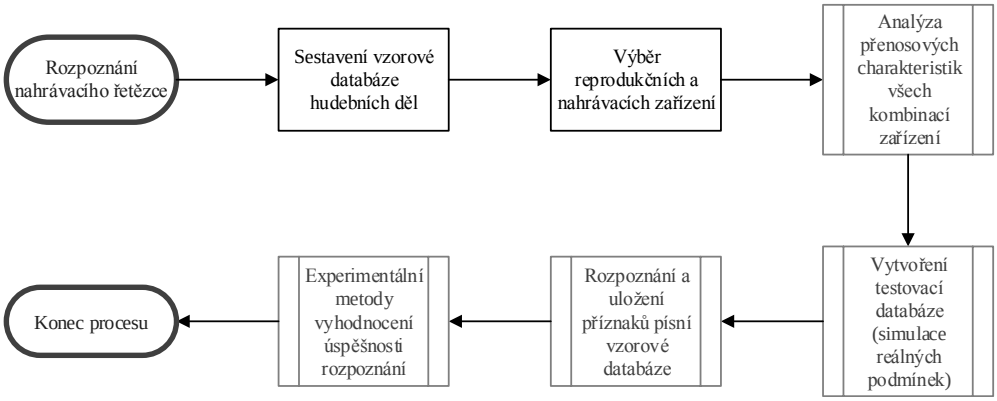
Pro simulaci různých nahrávacích prostředí byla do nahrávek při vytváření testovací databáze zakomponována impulzní odezva místností. To způsobilo, že nahrávka „byla přehrána“ právě v této místnosti a byly zvýrazněny její specifické vlastnosti. V reálu si to lze představit jako nahrávání ve školní učebně, koncertní síni, hospodě při zábavě s přáteli a dalších prostorách. Pokud tedy přenosové vlastnosti všech kombinací vybraných zařízení navíc přenesu do nahrávacích prostředí, získám další zobecnění testu. To je dostatečně obecné na to, aby mohl systém rozpoznávat písně a vyhodnocovat jejich nalezení v reálných podmínkách.

Vzorová databáze písní musí být redukována do souborů. Tyto soubory budou každou píseň charakterizovat. Jde totiž o to, aby se při každém pokusu o identifikaci nemusela celá databáze znovu parametrizovat. Soubory tedy budou obsahovat příznakové informace o každé dostupné písni v databázi. Na základě těchto informací je založena celá rozpoznávací architektura. Jak je patrné, příznaky každé písně musí být jednoznačné, aby byla zachována vzájemná odlišitelnost. Jen pro informaci, několika set gigabytová databáze písni je stlačena do příznakových vektorů, uložených v souborech o velikosti pár stovek megabajtů. Kompresní podíl má tedy hodnotu tří řádů.

Posledním, stěžejním krokem práce je vytvořit samotnou rozhodovací logiku. Ta by měla neznámou, hledanou píseň odlišit od ostatních nahrávek v databázi a

identifikovat ji. Tato logika je experimentální, čili obsahuje několik variant vyhodnocovacích metod. Každá varianta přináší svá pro i proti, ať už v podobě výpočetního výkonu nebo dalších. Tyto parametry lze uživatelsky upravovat, a tak docílit co nejoptimálnějšího řešení, v podobě přívětivého rozpoznávacího skóre.

Společnými vlastnostmi, na kterých rozpoznávací metriky a celá práce staví, je rozpoznávání tempa a začátků dob písně, ze kterých jsou vybrány frekvenční složky, specifikující dobu. Závěrem zbývá jen vyhodnocení navržených řešení a jejich vzájemné porovnání.



Obrázek 2.1: Celkový průběh návrhu rozpoznávacího systému

2.2 Vzorová databáze hudebních děl

2.2.1 Struktura hudební databáze

Pro splnění třetího bodu zadání bylo třeba vybrat databázi alespoň 600 děl. Shromáždil jsem výběry děl různých hudebních žánrů. Výběry obsahují skladby od 30. let minulého století, až po díla ze současné světové i domácí tvorby. Vybrané žánry byly roztříděny do 16 skupin, přičemž na každý žánr připadá 38 skladeb. Vzorová databáze bude tedy obsahovat počet skladeb rovný součinu těchto dvou údajů, celkem bude možné rozpoznat 608 děl.

Díla lze kromě hlavního dělení také roztřídit na jiné podskupiny. Příkladem může být rozdělení dle rychlosti skladby, čili jejího tempa. Dalším rozdělením pak může být hudba instrumentální a zpívaná, třídění dle pohlaví interpreta, dle jazyka zpěvu, a dalších vlastností.

Každý žánr má své charakteristické vlastnosti rytmu, používaných hudebních nástrojů, hudebního zabarvení, které ho odlišuje od ostatních. Na základě tohoto faktu jsem usoudil, že nejlepší bude hudbu třídit a vyhodnocovat rozpoznávací skóre mimo jiné právě dle žánru.

Žánry, jejichž seznam je uveden v tabulce 2.1, bych Vám teď chtěl alespoň rámcově přiblížit, aby bylo zřejmé, čím se navzájem odlišují, a tedy proč jsem se rozhodl sortovat databázi právě tímto způsobem. Ještě je třeba dodat, že zejména žánry,

blues	jazz
country	klasická hudba
dechová hudba	pop
disco	rhytm and blues
elektronická hudba	reggae
folk	rock
funk	ska
hip hop	soul

Tabulka 2.1: Seznam žánrů

které ve svém projevu nepoužívají výrazně bicí nebo je nepoužívají vůbec, výrazně zkomplikují funkčnost výše zmíněných součástí, jako je rozpoznání tempa a začátků dob písně. Další možnou komplikací je žánr jazz, kde se s oblibou používá tzv. triolový feeling. To znamená, že kolísá délka osminových not.

2.2.2 Charakteristika jednotlivých žánrů

Blues vznikl v afro-americké otrocké komunitě v 19. století. Je inspirován světskou hudbou, je také ovlivněna africkou hudbou původních národů. Slovo blues znamená deprese, smutek, proto je těmito slovy inspirováno velké množství skladech tohoto žánru. Veselé a rychlejší skladby však nejsou výjimkou, typický znak je však houpaný rytmus skladeb a používá synkop (úmyslné předražení doby), dále potom hudební nástroje, například kytara, klavír, harmonika, saxofon, trubka a pozoun, bicí. Legendami jsou Eric Clapton a Ray Charles.

Country je původem americký hudební styl, který vznikl koncem 18. století. Jeho hlavní sláva však přišla v první polovině 20. století. Na americkém a kanadském venkově je velmi populární i dnes. Inspiraci nalézá v hudbě španělských, francouzských a irských přistěhovalců. Písně jsou většinou rychlejší, mají tematicky užší okruh, jako láska, venkov, farmaření, cestování. Typickými nástroji jsou kytara, banjo, housle, harmonika, mandolína a bicí. Slavnými interprety jsou Johny Cash a Tim McGraw.

Dechová hudba nalezla svůj původ v evropské vážné hudbě a folklóru. Vyvinul se začátkem 19. století. Písně mají především pochodový a taneční nádech, v České republice je oblíbený především na východě. Pro svůj projev využívá především žesťové nástroje a perkuse. Hrdým zástupcem je soubor Moravanka.

Disco je žánr taneční hudby, který se vznikl v afro-americké a hispánské komunitě koncem 70. let 20. století. Vyvinul se především z funku a Latinskoamerické hudby. Charakterizuje ho šestnáctinový či osminový rytmus prokládaný synkopami jiných nástrojů. Má bohaté využití hudebních nástrojů, jak dechových, tak i elektrických. Nejznámějšími představiteli jsou ABBA či Boney M.

Elektronická hudba je termín pro hudbu vytvářenou pouze elektronickými součástmi, jako teremín, syntetizér, sampler. Elektromechanické nástroje, kterým je například elektrická kytara, do této skupiny nepatří. Z toho plyne, že bude mít velmi přesné tempo.

Folk je hudební žánr mající své kořeny v anglosaských zemích. Folk znamená lid, čili je to lidový žánr. Typickými nástroji jsou akordeon, harmonika, mandolína, akustická kytara a bubny. Světoznámými představiteli tohoto směru jsou Bob Dylan a Bruce Springsteen.

Funk byl vytvořen afroameričany v 60. letech minulého století. Struktura skladby bývá jednodušší, vyznačuje se čtyřčtvrtečním taktem, silnou basovou linkou a ostrou rytmickou kytarou. Důraz se klade i na taneční stránkou hudby. Nejznámějšími představiteli jsou James Brown a Kool & The Gang.

Hip Hop je název kultury, která vznikla počátkem 70. let 20. století mezi lidmi na okraji společnosti. Kořeny tohoto stylu sahají až na Jamajku. Charakteristický jev je, že hudba je vytvářena pomocí hudebních mixů, prostřednictvím dvou gramofonů a mixážního pultu. Vokální projev je postaven na rytmizaci jazyka, hrou se slovy a beatboxingu, neboli napodobování bicích nástrojů za pomoci lidského hlasu.

Jazz vznikal začátkem minulého století mezi afroamerickou komunitou na jihu Spojených států amerických smísením afrických a evropských hudebních stylů. Vyznačuje se používáním synkop, neznělých tónů a improvizaci. Používanými hudebními nástroji jsou především saxofon, trubka, klarinet, pozoun, kytara, Hammondovy varhany, klavír, kontrabas a bicí. Nejslavnějším propagátorem tohoto žánru je bezpochyby Louis Armstrong.

Do klasické hudby nebo také vážné hudby se řadí zejména evropská hudební tradice z období renesance, baroka, klasicismu a romantismu. Je charakteristická pouze pro západní kulturu. Každé období mělo své charakteristické hudební rysy, ale všeobecně se kladl důraz na harmonii a prokládání rychlých části s volnými. Typickými nástroji jsou housle, viola, violoncello, kontrabas, varhany, cembalo a především klavír. Jelikož se v tomto hudebním směru nepoužívají bubny, chybí zde rytmika. Nejobdivovanější skladatelé své doby byli Ludwig van Beethoven, Antonio Vivaldi, Carl Philip Emmanuel Bach, Frédéric Chopin a Antonín Leopold Dvořák.

Pop vznikl v 60. letech 20. století v USA. Vyznačuje se výraznou zpívanou melodií, doprovázenou moderním způsobem. Cílem skladby je zaujmout co nejvíce posluchačů a stát se komerčně úspěšnou. Sloky a refrén mají předvídatelnou strukturu. Používanými nástroji jsou zpravidla klávesy, akustická kytara, piano, syntezátor, sekvencer. Nejznámějšími interprety jsou Billy Joel, Michael Jackson, Elton John a Madonna.

Rhytm and Blues je hudební žánr populární hudby vytvořený Afroameričany. Vznikl v 40. letech minulého století a kombinoval prvky jazzu, gospelu a blues. V současné době začal, podobně jako jiné styly, přejímat prvky pop rocku. Nejčastějšími nástroji původního R&B byly basová kytara, saxofon, bicí a klávesové nástroje.

Reggae hudební styl, který se vyvíjel v průběhu 60. let na Jamajce. Texty obsahovaly většinou zábavná témata, občas se projevila i sociální kritika na Jamajce. Typická rastafariánská tématika se objevila až v 70. letech. Typickými nástroji jsou baskytara, kytara, elektrické varhany, žesťové nástroje. Jména jako Bob Marley, Damian Marley nebo Rastafari jsou známá nejen v reggae komunitě.

Rock je žánr populární hudby, jehož i dodnes trvající sláva začala v 60. letech 20. století. Má kořeny v rock and rollu, country, ale také folku. Nejčastěji používanými nástroji jsou elektrická kytara, baskytara, bicí a klávesové nástroje, ale také

Účel	Typ	Název	Varianta
Reprodukce	Notebook	Asus K50ID	-
Reprodukce	Projektor	BenQ MS-500H	-
Reprodukce	2.1 reproduktory	Creative	se subwooferem
Reprodukce	2.1 reproduktory	Creative	bez subwooferu
Reprodukce	2.1 reproduktory	GX Gaming SW-G2.1 1250	se subwooferem
Reprodukce	2.1 reproduktory	GX Gaming SW-G2.1 1250	bez subwooferu
Reprodukce	Mobilní telefon	Samsung GSH-i900	-
Nahrávání	Mobilní telefon	Apple iPhone 4	-
Nahrávání	Náhlavní sluchátka	GemBird HeadPhones	-
Nahrávání	Notebook	HP Probook 4530s	-

Tabulka 2.2: Kombinace analyzovaných zařízení

harmonika. Slavných představitelů je spousta, pro příklad lze uvést skupinu Queen, Aerosmith nebo Kiss.

Ska je hudební styl pocházející z Jamajky, kde se vyvinul koncem 50. let. Je považován za předchůdce reggae hudby. Má typický kolísavý rytmus, je však rychlejší než reggae a často se mísí s punkem nebo jazzem. Typickými nástroji jsou kytara, baskytara, trubka, pozoun, saxofon, piano, bicí a varhany. Nejznámějšími představiteli jsou Inner Circle či němečtí Irie Révoltés.

Soulová hudba vznikla z afroamerického gospelu. Jejím základem byl spirituálový hymnus s bohatou harmonií a rytmem. Používanými nástroji jsou saxofon, klavír, dechové nástroje a varhany. Průkopníkem soulu byl Ray Charles, do obliby se dostal především díky Jamesi Brownovi.

2.3 Nahrávací zařízení

2.3.1 Výběr zařízení

Pro zachování objektivity při navrhování řešení systému bylo třeba brát v potaz přenosové vlastnosti zařízení. Například zařízení určená k reprodukci, jako reprobedny, budou mít lepší přenosové vlastnosti, než reproduktor mobilního telefonu, který je určen především pro přenos frekvencí odpovídajících hlasu. To znamená, že basy a nižší tóny vůbec nepřenese. Vybral jsem proto ty z nejčastějších, s jakými se lze v běžném životě setkat.

2.3.2 Kombinace zařízení

Zařízení jsem rozdělil do dvou skupin. Jednou skupinou jsou zařízení reprodukční, druhou pak nahrávací. Tabulka 2.2 přesně zobrazuje typy a názvy zařízení, která jsem použil.

Jak lze z tabulky vyčíst, bylo použito celkem pět druhů reprodukčních zařízení. V případě 2.1 reprosoustav bylo dělení ještě rozdvojeno na případ použití při zapnuté/vypnuté sub-basové jednotce neboli subwooferu. Celkem jsem tedy pracoval se sedmi nahrávacími sestavami. Dále byly vybrány 3 druhy zařízení nahrávacích.

2.4 Měření přenosových charakteristik zařízení

2.4.1 Nahrávání

Nahrávání probíhalo proto, abych následně mohl vyjádřit přenosové charakteristiky testovacích zařízení. To bylo důležité k tomu, abych následně mohl simulovat přehrávání skladeb na jednotlivých kombinacích zařízení a vytvořit testovací databázi, která simuluje reálné nahrávací podmínky.

2.4.2 Program Audacity

Audacity [1] je multiplatformní editor digitálního zvuku. Byl vytvořen Dominikem Mazzonim, který v současné době pracuje ve společnosti Google. Zdrojový kód je uvolněn pod licencí GNU.

Překypuje celou škálou funkcí, jako převod nahrávek z pásků a desek do digitálního záznamu, editace zvukových souborů různých formátů, stříhání, rozdělování a míchání nahrávek, odstranění šumu a další.

Pro mou potřebu bylo nutné, aby program umožňoval pomocí tónového generátoru vytvořit a uložit audio signál v určitém rozmezí frekvencí, které se v čase mění. Tuto vlastnost program Audacity splňuje, proto jsem si ho vybral jako generátor signálu pro měření přenosových charakteristik zařízení.

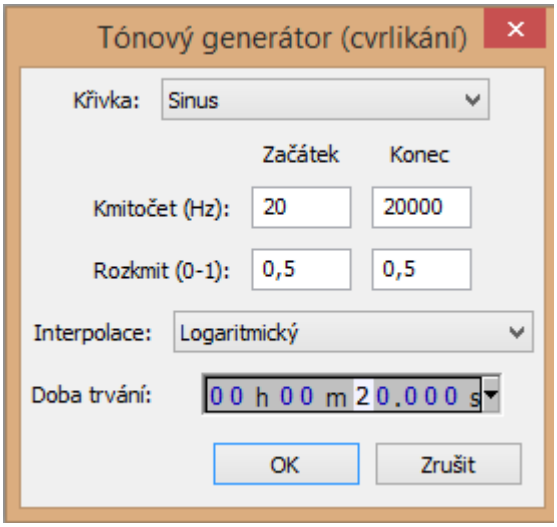
2.4.3 Signál slyšitelného spektra

Při vytváření tohoto signálu v programu Audacity lze nastavit, s jakými vlastnostmi je signál vygenerován 2.2. Jsou jimi typ křivky, kde lze vybrat sinusový, čtvercový či trojúhelníkový průběh signálu. Dále kmitočet, kde vybírám rozsah generovaných frekvencí a rozkmit, kde se nastavuje amplituda signálu. Následujícím parametrem je interpolace, kde se nastavuje, zda se budou frekvence měnit lineárním nebo logaritmickým průběhem. Poslední vlastností je doba trvání signálu.

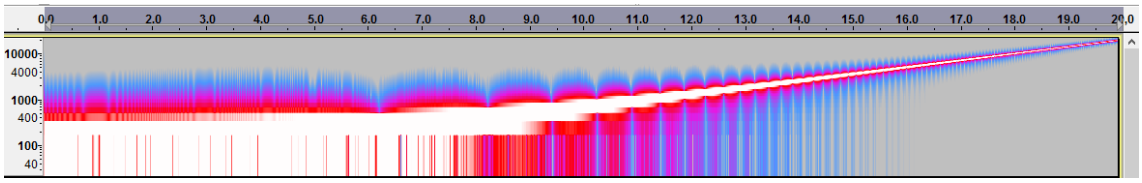
Tento signál jsem generoval se sinusovým průběhem, který pro lidské ucho zní jako plynulý. Kmitočet jsem nastavil jako zvyšující se rozsah slyšitelného frekvenčního spektra, čili 20 Hz až 20 kHz. Lidský sluch vnímá zvuk logaritmicky, dle toho byl také nastaven parametr interpolace. Doba trvání generovaného signálu je 20 sekund. Jeho spektrogram si lze prohlédnout na obrázku 2.3.

Jelikož jsem přenosové vlastnosti měřil pro tři nahrávané hlasitosti, je poslední parametr, rozkmit, proměnný.

[1] Web: audacity.sourceforge.net



Obrázek 2.2: Nastavení tónového generátoru



Obrázek 2.3: Spektrogram vygenerovaného signálu s logaritmickým průběhem

2.4.4 Parametry nahrávání

Nahrávání přenosových vlastností zařízení probíhalo v místnosti o ploše cca. 14 m^2. Vzdálenost nahrávacího zařízení od reprodukčního byla jednotná, a to přibližně 1 metr. Snažil jsem se o to, aby bylo každé reprodukční zařízení nastaveno na 85% svého výkonu. Tato parametrizace zajistila, aby se vlastnosti napříč zařízeními projevily rovnoměrně. Nahrávání probíhalo ve 3 hlasitostech, abych zjistil, jak se přenosové vlastnosti zařízení mění na základě jiné nahrávané hlasitosti. Amplituda generovaného signálu, která představovala nahrávanou hlasitost, byla 0,2 pro nejtišší, 0,5 pro střední a 0,9 pro hlasitou nahrávku. Vzorkovací frekvence generovaného signálu byla 44,1 kHz.

2.4.5 Průběh nahrávání

Záznam nahrávek probíhal následujícím způsobem. Nejprve byla obě zařízení umístěna do stanovených poloh. Na reprodukčním zařízení byla spuštěna nejtišší vytvořená nahrávka. Ta byla zaznamenána a uložena na všech nahrávacích zařízeních zároveň. Všechna nahrávací zařízení mají vestavěný pouze jeden mikrofon, záznam tedy probíhal v režimu mono. Tento postup se opakoval pro každou generovanou hlasitost nahrávky. Poté bylo reprodukční zařízení vyměněno a celý postup

byl zopakován. Po záznamu všech potřebných stop a jejich přenosu do počítače jsem mohl přistoupit k analýze přenosových vlastností zařízení.

2.4.6 Úprava nahrávek

Přenesené nahrávky musely být před analýzou ještě upraveny. Je totiž zřejmé, že délka nahrávky nemůže být stejná, jako délka reprodukovaného originálu. Jinými slovy, na nahrávacím zařízení jsem nemohl stlačit nahrávací tlačítko ve stejný okamžik, jako přehrávací tlačítko na zařízení reprodukčním. Skutečnost je taková, že byl nahrávací symbol stlačen o něco dříve.

Po zastavení nahrávání jsem ručně vyřízl a uložil takovou část nahrávky, která odpovídala experimentu. Ořízl jsem tedy takovou část nahrávky, která evidentně odpovídala šumu v okolí.

2.4.7 Analýza nahrávek

Nahrávky všech kombinací zařízení byly analyzovány pomocí programu Matlab. Bylo potřeba zjistit, jaké frekvence jednotlivá zařízení přenesou, jak je zesílí či utlumí. Všechny údaje bylo potřeba analyzovat zvlášť pro každou nahrávací úroveň hlasitosti. Spočtené charakteristiky byly následně také zobrazeny v grafech. Celý postup zachycuje vývojový diagram 2.4 (funkce **FrequencyGainMeasurement**).

Prvním krokem bylo tedy načíst cesty k nahrávkám uloženým v počítači do matice. Ta slouží pro načítání adresářové struktury.

Dále je třeba vypočítat průběh přenosových charakteristik (viz diagram 2.5 (funkce **evaluateGain**)) pro danou nahrávací kombinaci, pro každou analyzovanou hlasitost.

Po načtení nahrávky do vektoru je signál doplněn na délku dvaceti sekund, tedy na délku původního originálu (důvody viz podkapitola s názvem Úprava nahrávek)

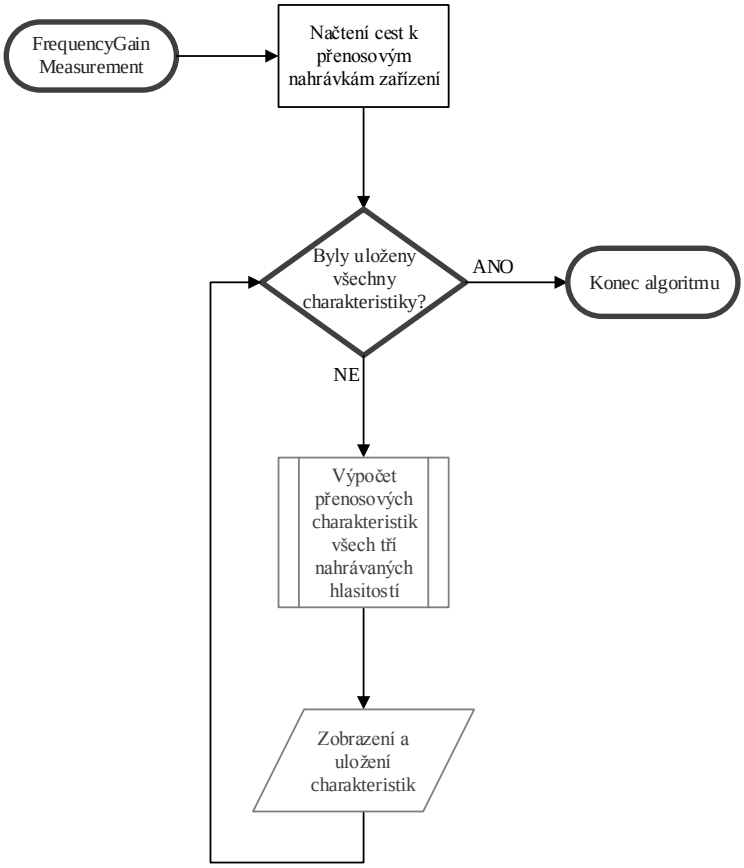
Dále bylo třeba vyjádřit zesílení každé vlny signálu. Procházím tedy celý signál a hledám lokální maximum a minimum. Pokud byly tyto hodnoty nalezeny, uložím jejich průměrnou absolutní hodnotu do proměnné popisující zesílení. Tato hodnota vyjadřuje průměrné zesílení vlny. Poté se v analýze signálu přesunu na další vlnu.

Poté, co jsem tímto způsobem analyzoval všechny tři nahrávané hlasitosti, překročil jsem k zobrazení vyjádřených charakteristik. Každá zobrazená charakteristika byla následně uložena.

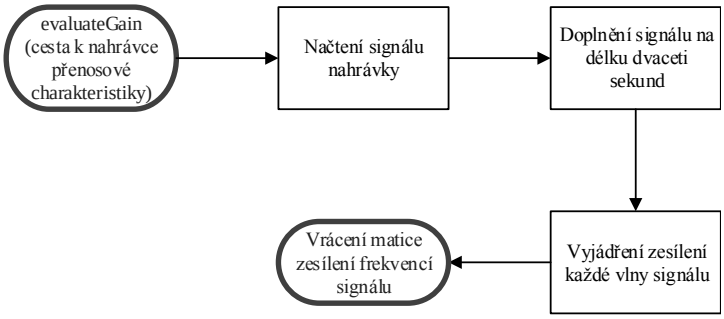
Přenosové charakteristiky kombinací zařízení jsou uloženy na přiloženém DVD.

Pro vizuálně přívětivou reprezentaci přenosových charakteristik nebyl tento způsob zobrazení vhodný. Bylo sice poznat, jaká byla amplituda vzhledem k originálnímu signálu, avšak údaje nebyly příliš čitelné.

Dalším krokem tedy bylo zobrazit tyto data lepším, čitelnějším způsobem. Vzhledem k tomu, že většinovou náplní práce je analýza hudebních nahrávek, bylo vhodné uchýlit se k zobrazení přenosových charakteristik zařízení dle frekvencí not. Z toho způsobu zobrazení již bude možné s jistotou usoudit, jak jednotlivé kombinace zařízení zesílí či utlumí tóny na vstupu.



Obrázek 2.4: Průběh při analýze přenosových charakteristik zařízení



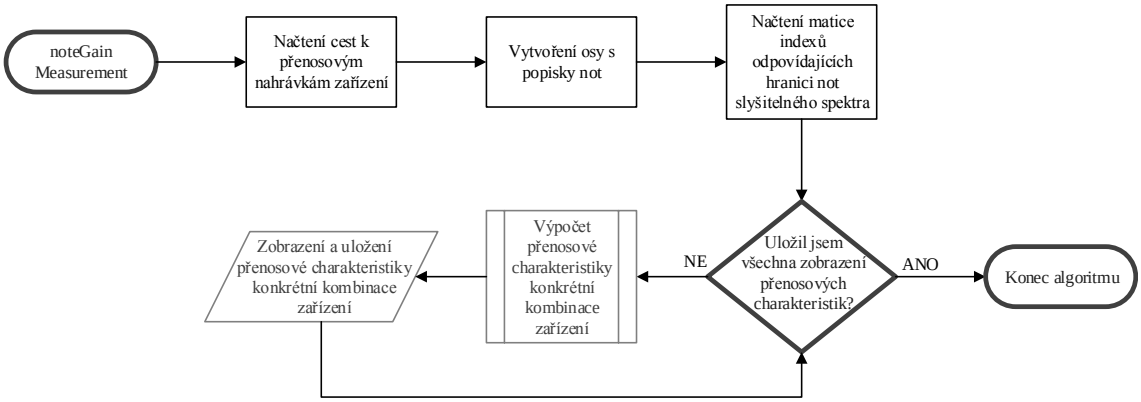
Obrázek 2.5: Průběh při výpočtu amplitudy signálu

2.4.8 Přepočet a zobrazení na notové ose

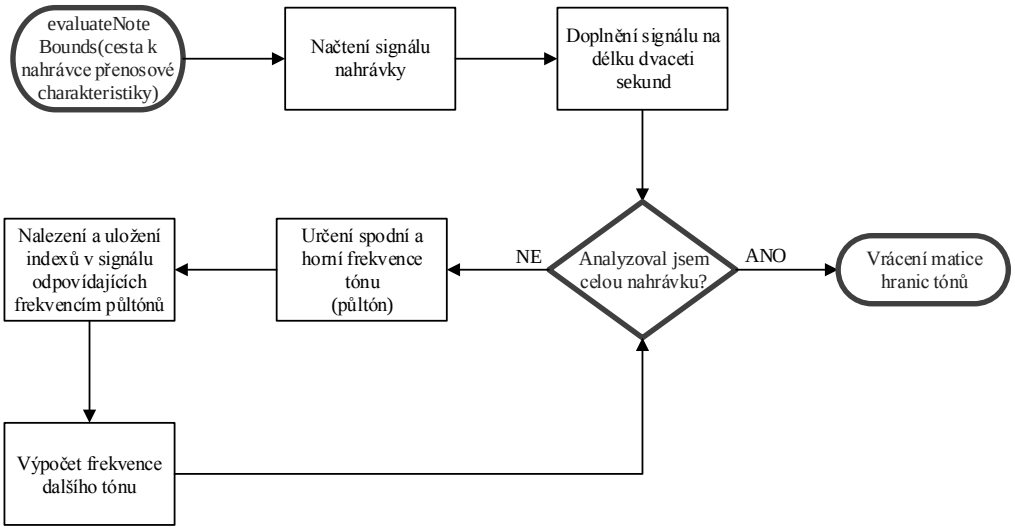
Obecný postup pro tento krok je zobrazen na vývojovém diagramu 2.6 (funkce **noteGainMeasurement**).

Po načtení cest k přenosovým nahrávkám a vytvoření osy pro zobrazení popisků not jsem potřeboval pro každou z nich vyjádřit, v jakém čase zazněla frekvence tónů, které chci zachytit. Stěžejní kroky funkce jsou zachyceny na diagramu 2.7 (funkce

evaluateNoteBounds).



Obrázek 2.6: Průběh při přepočtu přenosových charakteristik na notovou osu



Obrázek 2.7: Vyjádření hranic tónů v čase

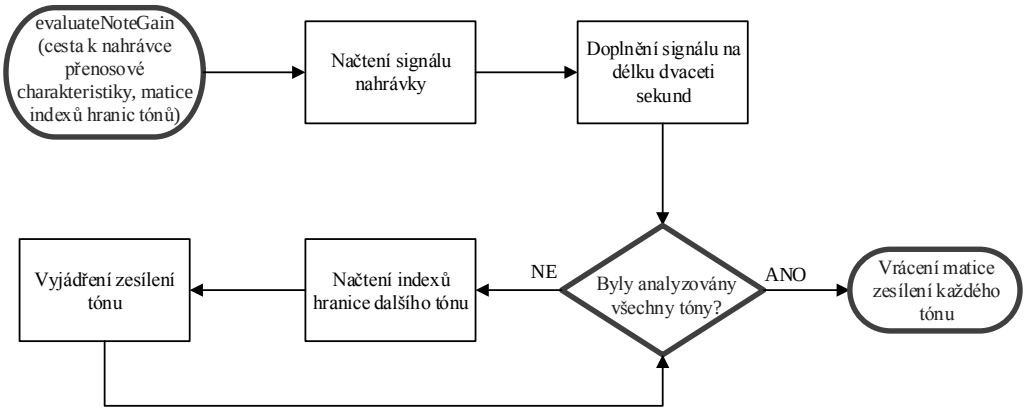
Poté jsem mohl přikročit k načtení signálu nahrávky a doplnění na délku 20 sekund, obdobně jako u funkce 2.4, jsem mohl přistoupit k samotnému nalezení indexů. To funguje tak, že pro každou frekvenci tónu jsem spočítal její spodní a horní hranici (půltón). Výpočet je realizován pomocí vztahu 2.1.

$$f_2 = \sqrt[24]{2} f_1 \tag{2.1}$$

Kde:
f_2 je frekvence půltónu [Hz],
f_1 je frekvence tónu [Hz].

Frekvence byla za předpokladu, že vzorkovací frekvence byla známa, přepočítána na periodu mezi vlnami, kterou jsem následně při průchodu signálem hledal. Nalezené indexy začátků byly uloženy do výsledné matice a bylo přistoupeno k nalezení indexů dalšího tónu. Postup byl opakován do doby, než byl zpracován celý signál. Výstupem z této funkce je seznam indexů, odpovídající začátkům znění tónů v čase. Vzhledem k faktu, že všechny nahrávky mají v čase stejné spektrum frekvencí, stačilo provést tuto analýzu pouze pro jednu nahrávku. Výstup funkce byl poté pro zvýšení výkonu systému a další uplatnění uložen. Blok kódu byl nahrazen načtením proměnné ze souboru.

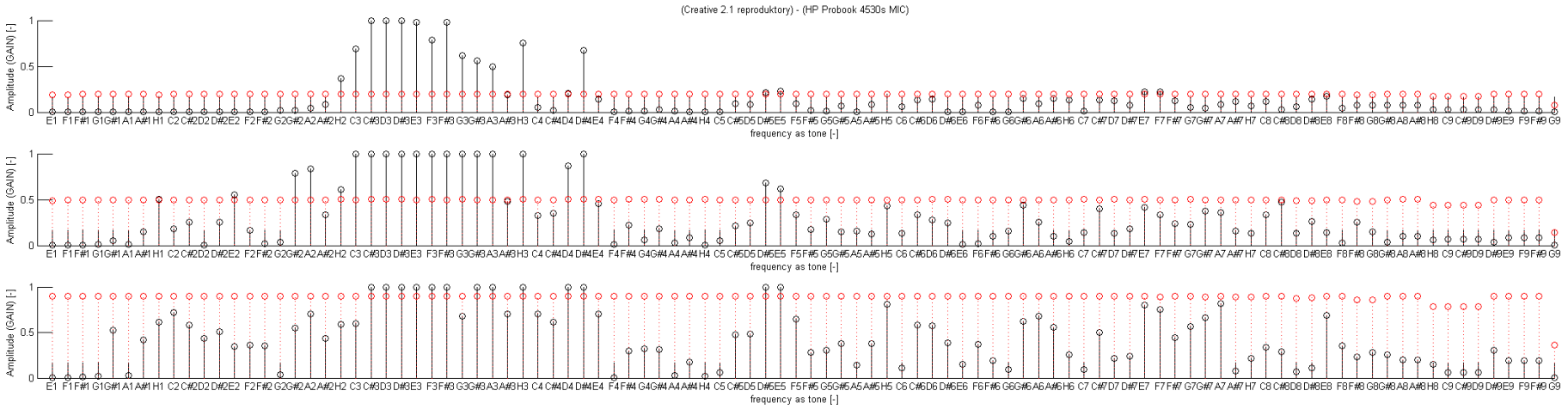
Poté jsem mohl přejít k samotnému výpočtu zesílení ve vztahu k jednotlivým tónům. Princip tohoto kroku opět vysvětluje vývojový diagram 2.8 (funkce **evaluateNoteGain**).



Obrázek 2.8: Výpočet zesílení jednotlivých tónů

Jak lze z diagramu vyčíst, do funkce vstupuje matice obsahující indexy hranic tónů. První dva kroky postupu jsou shodné s předchozí funkcí. Poté postupně procházím signál, který je pomocí indexů rozsegmentován na části, z nichž každá obsahuje frekvence jednoho tónu. Pro každou tuto část bylo vypočítáno průměrné zesílení. Výpočet probíhal pomocí vyjádření průměrného zesílení každé vlny (opět pomocí lokálních maxim a minim), které bylo následně uloženo. Celkové průměrné zesílení tónu bylo vyjádřeno jako suma průměrných zesílení dělená počtem vln v segmentu. Po analýze každého segmentu byla vrácena matice průměrných zesílení tónů. Ta byla pro další použití uložena.

Nyní se přistoupilo k poslední části tohoto algoritmu, a to samotnému zobrazení a uložení přenosových charakteristik. Tento proces je analogicky obdobná sekvence příkazů, jako při zobrazování amplitudy v čase. Výsledný graf si lze prohlédnout na obrázku 2.9.



Obrázek 2.9: Zesílení jednotlivých tónů

Přenosové charakteristiky dalších kombinací zařízení z pohledu zesílení tónů jsou uloženy na přiloženém DVD.

2.4.9 Zhodnocení a rozhodnutí pro další postup

Po důkladném prostudování všech uložených charakteristik jsem došel k následujícím, logickým a předpokládaným závěrům, které shrnují obrázky 2.10 a 2.11.

			Reprodukční zařízení			
			Notebook	*Projektor*	*2.1 reproduktory*	
			Asus K50ID	*BenQ MS-500H*	*Creative*	*Creative*
			-	-	*se subwooferem*	*bez subwooferu*
Nahr. z.	*Mobilní telefon*	*Apple iPhone 4*	G5 - F9	A5 - C#7	D4 - G#8	C5 - D9
	Náhlavní sluchátka	*GemBird HeadPhones*	C6 - D9	G5 - H6	G2 - G#8	C3 - F9
	Notebook	*HP Probook 4530s*	C5 - D9	C#5 - F#8	H1 - F#9	A#2 - D#9

Obrázek 2.10: Přenosová pásma analyzovaných zařízení (1)

			Reprodukční zařízení		
			2.1 reproduktory		*Mobilní telefon*
			GX Gaming SW-G2.1 1250	*GX Gaming SW-G2.1 1250*	*Samsung GSH-i900*
			se subwooferem	*bez subwooferu*	-
Nahr. z.	*Mobilní telefon*	*Apple iPhone 4*	C4 - D#9	H4 - F#9	E6 - C9
	Náhlavní sluchátka	*GemBird HeadPhones*	G2 - H8	C5 - A7	G6 - H8
	Notebook	*HP Probook 4530s*	F#2 - G#8	D#4 - A8	C6 - D9

Obrázek 2.11: Přenosová pásma analyzovaných zařízení (2)

Je patrné, že mikrofony mobilních telefonů a obyčejných náhlavních sluchátek, konstruované především pro přenos lidského hlasu, přenesou frekvence mezi 1-3 kHz. Mikrofon notebooku měl nahrávací spektrum o něco širší.

Ke zhodnocení reprodukčních zařízení lze říct, že reprosoustavy 2.1 s aktivní basovou jednotkou přenesly znatelně větší část frekvenčního spektra. Z grafů lze vyčíst, že nízké frekvence, odpovídající basům, byly při středním a hlasitém nahrávání vzhledem k originálnímu signálu přebuzeny. Reproduktor projektoru je spíše doplňkem. Je znát, že konstruktéři spoléhají na připojení externího zvukového zdroje.

Dalším pozorovaným faktem bylo, že zesílení mezi dvěma sousedními notami v některých případech značně kolísá. Tuto nepříjemnost, jejíž možná příčina je dána rezonací krytu telefonu či volných předmětů v místnosti (například lžíce nebo sklenice), se snažím odstranit v další kapitole mé práce.

Nakonec ještě bylo třeba stanovit kompromis frekvencí, které rámcově charakterizují všechna analyzovaná zařízení, tedy pokryjí jejich přenosové vlastnosti. Vybrané rámcové frekvence jsou 51.913 Hz jako spodní mez a 6271.928 Hz jako mez horní. Tyto frekvence poté využívám k vytvoření filtru typu „pásmová propusť“, který uplatňuji při zjišťování tempa skladby.

2.5 Simulace reálných podmínek, vytvoření testovací databáze

Další částí práce bylo přenést naměřené charakteristiky do nahrávek a simulovat tak realitu. Postup při návrhu filtrů opět zachycuje vývojový diagram 2.12 (funkce **signalStabilization**).

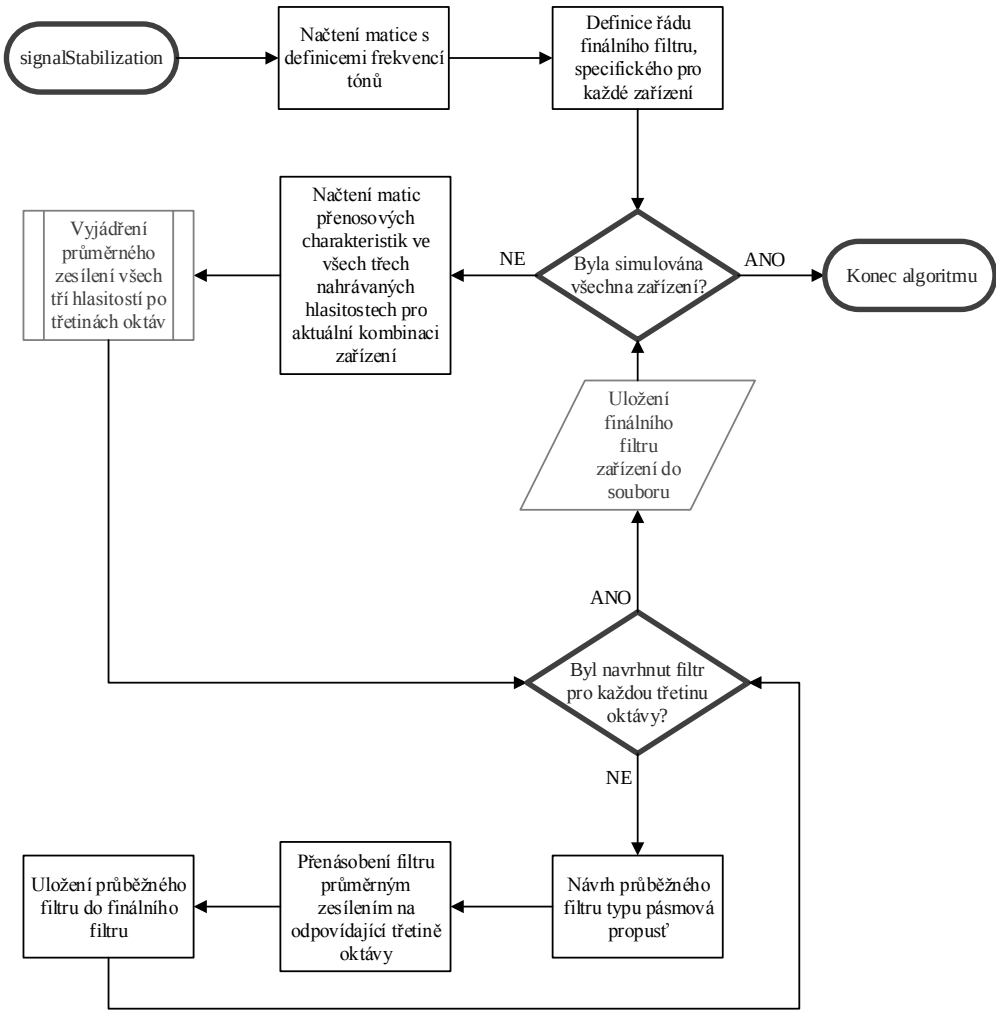
Realita je specifikována testovací databází, která obsahuje každou skladbu originální databáze. Skladby byly přefiltrovány, aby bylo simulováno přehrání na jednotlivých kombinacích zařízení. Navíc bylo simulováno reálné nahrávací prostředí, a to pomocí impulzních odezev tří místností pro každou skladbu. Nyní budou rozebrány jednotlivé části podrobněji.

2.5.1 Definice parametrů filtrů

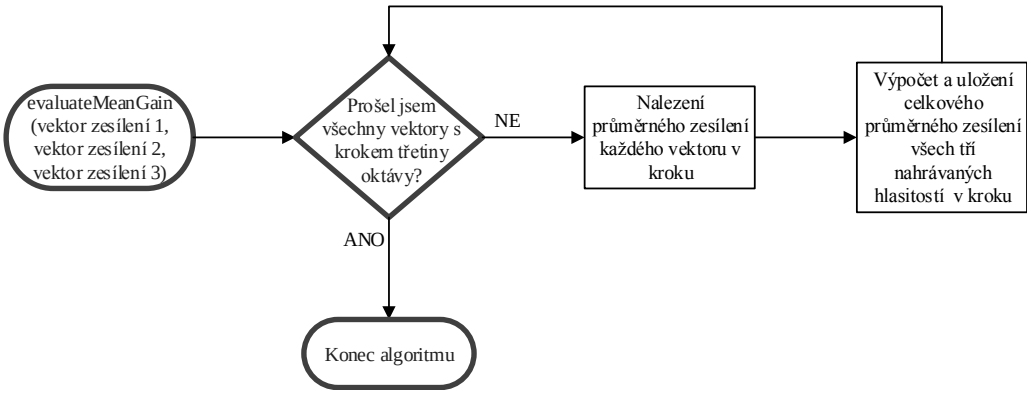
Prvním krokem bylo načíst matici frekvencí tónů. Její obsah jsem použil pro definici mezních frekvencí při vytváření pásmové propusti.

Poté bylo třeba definovat parametry filtrů. Díky tomu, že zesílení mezi dvěma sousedními notami v některých případech kolísalo, rozhodl jsem se signál stabilizovat a filtrovat pomocí průměrného zesílení po třetinách oktáv. Potřeboval jsem proto přesnější filtr, aby zlom filtru třetiny oktávy u nízkých tónů nepřesahoval do filtru oktávy vyšší. Řád filtru byl proto zvolen na hodnotu 1000.

Ze souborového systému se načetly matice průměrných zesílení tónů pro všechny tři nahrávané hlasitosti pro konkrétní kombinaci zařízení. Poté bylo třeba vyjádřit průměrné zesílení po třetině každé oktávy. Tento postup popisuje vývojový diagram 2.13 (funkce **evaluateMeanGain**).



Obrázek 2.12: Postup při stabilizaci signálu

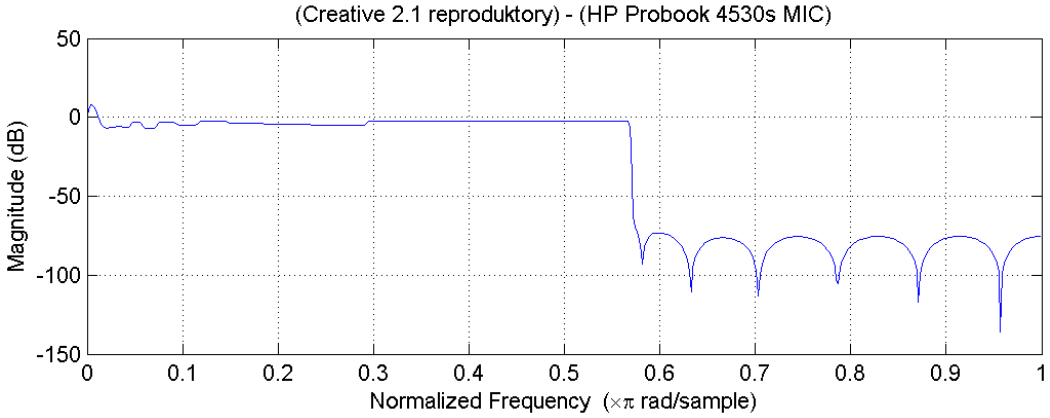


Obrázek 2.13: Postup při výpočtu průměrného zesílení

Procházel jsem matici a zjišťoval, který tón měl maximální zesílení na třetině analyzované oktávy pro každou nahrávací hlasitost. Poté bylo třeba z těchto maxim udělat relativní veličinu, a to proto, abych mohl zesílení vzájemně porovnávat mezi nahrávacími hlasitostmi. Následně byla maxima ještě jednou zprůměrována, a to napříč hlasitostmi. Dostal jsem tedy jedno číslo, charakterizující průměrné zesílení na třetině oktávy. Postup jsem opakoval, dokud nebyly analyzovány všechny tóny. Do hlavního programu byla vrácena matice průměrných zesílení.

2.5.2 Návrh filtrů

Dalším krokem již mohl započít samotný návrh filtrů. Filtry jsem analogicky navrhoval také po třetinách oktáv. Po stanovení hraničních frekvencí pásmové propusti byl samotný průběžný filtr vynásoben právě průměrným zesílením, korespondujícím s navrhovaným filtrem. Tento filtr byl poté přičten do finálního filtru. Po analýze všech třetin byla frekvenční odezva finálního filtru (viz obrázek 2.14) vykreslena a uložena.



Obrázek 2.14: Finální filtr zařízení

Filtry ostatních zařízení jsou na přiloženém DVD. Filtr byl také uložen, protože byl využit při vytváření testovací databáze. Filtry jsem tvořil pro každou kombinaci zařízení.

2.5.3 Testovací databáze

Testovací databáze je soubor skladeb, jejichž věcný obsah (samotná hudba), byl obohacen o reálné nahrávací podmínky. Na originální databázi bylo tedy třeba aplikovat navržené filtry a impulzní odezvy místností.

Databáze má následující parametry. Počet obsažených skladeb ve vzorové databázi se vynásobí počtem nahrávacích kombinací (počtem filtrů) a počtem impulzních odezev místností, tedy 608×21×3. Obsahuje tedy 38304 variant skladeb pro testování. Obdobně jako u vzorové databáze, byla uložena ve formátu wav. Její velikost na disku po vytvoření je 789 GB.

Před samotným vytvářením této databáze byla třeba sehnat impulzní odezvy tří místností. Podmínkou bylo, aby se tyto tři místnosti navzájem odlišovaly odezvou (akustickými vlatnostmi). Z volně dostupných zdrojů [2] [3] [4] jsem tedy stáhl impulzní odezvu školní učebny univerzity v Londýně, která charakterizovala místnost s nejméně výrazným echem. Další místností byla Opera House v Sydney, tu jsem označil za místnost se střední dobou dozvuku. Poslední místností byla koupelna v Schulz Building ve městě Adelaide (Austrálie), která je střediskem výuky hudby. Ta v mé práci simulovala místnost s nejvýraznějším dobou dozvuku.

2.5.4 Způsob vytváření testovací databáze

Kompletní proces vytváření testovací databáze vystihuje diagram 2.15 (funkce **databaseCreation**).

Prvním krokem bylo načtení souborových cest ke skladbám originální databáze, dále filtrům simulující zařízení a impulzním odezvám místností. Poté jsem mohl přistoupit ke konkrétní modifikaci skladeb. Po načtení signálu hudební nahrávky z databáze se ve druhé smyčce postupně načítaly filtry konkrétních kombinací zařízení. Třetí smyčka sloužila pro postupné načítání impulzních odezev místností. Po načtení všech potřebných informací byla skladba profiltrována a byla simulována impulzní odezva místnosti. Ta byla do skladby aplikována pomocí metody Overlap-Add.

Metoda Overlap-Add ke své funkčnosti využívá Rychlou Fourierovu transformaci. Jak je známo, tento algoritmus je optimalizován pro signály, jejichž počet vzorků je roven mocnině dvou. Proto byla nejprve vyjádřena optimální délka pro výpočet FFT. Poté byly oba kanály hudebního signálu sloučeny do jednoho. Stejný krok byl proveden také u signálu impulzní odezvy místnosti. Následně byl na impulzní odezvu místnosti aplikován algoritmus FFT. Signál hudební nahrávky jsem procházel po částech, jejichž délka byla dána dříve vypočteným krokem. Z výřezu hudebního signálu je také provedena FFT. Dále byly oba signály pronásobeny po prvcích a byla provedena jejich zpětná (inverzní) transformace (IFFT), zpět do časové oblasti. Tato část byla uložena jako součást výsledného signálu s aplikovanou impulzní odezvou místnosti. Pro modifikaci všech částí byl finální signál nejprve vyvážen. To znamená, že jeho amplituda (hlasitost) byla převedena na rozsah 0-1. Poté byl signál vrácen zpět z funkce.

Posledním krokem bylo uložení modifikované skladby do souboru testovací databáze. Celý proces filtrací a aplikace impulzních odezev byl prováděn pro každou skladbu v originální databázi.

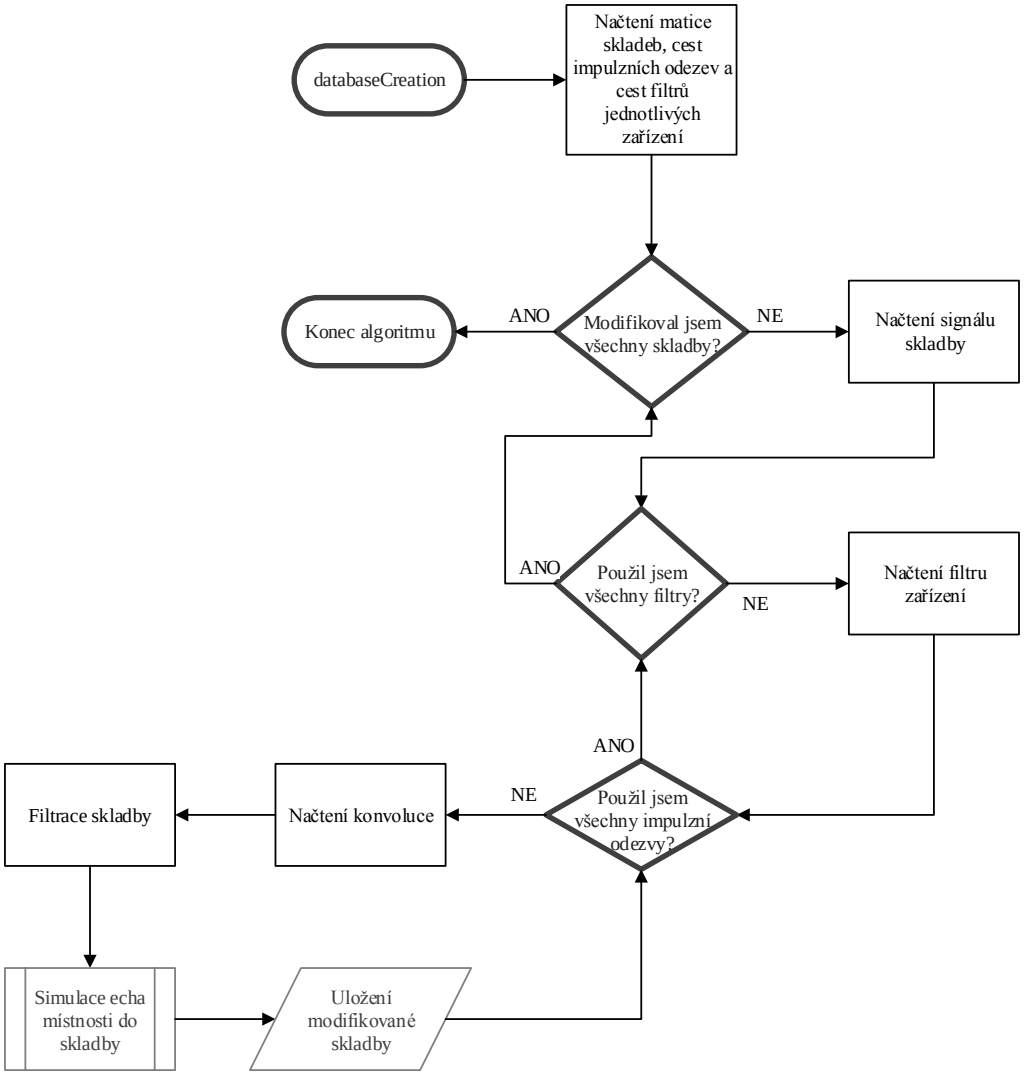
2.6 Rozpoznání příznaků děl

Další stěžejní částí práce byla samotná klasifikace děl. Bylo tedy třeba stanovit příznaky, které budou skladby charakterizovat. Na jejich základě bude poté možné

[2] isophonics.net/content/room-impulse-response-data-set

[3] little-scale.blogspot.cz/2012/10/impulse-response-pack-schulz-building_4413.html

[4] www.ee.usyd.edu.au/carlab/UserFiles/SOH/index.html



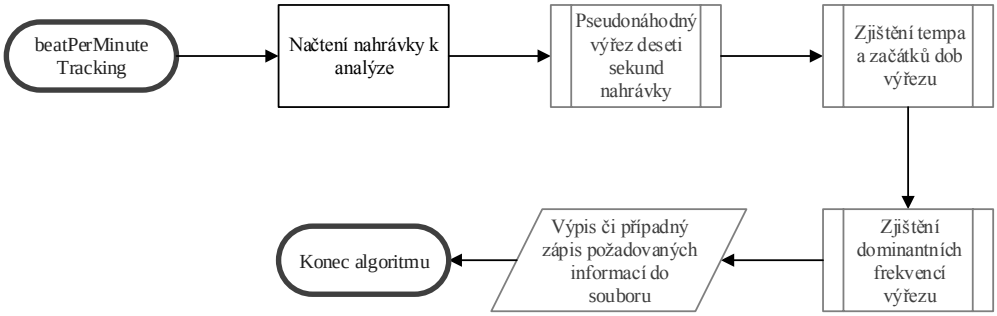
Obrázek 2.15: Postup při vytvážení testovací databáze

jednotlivé nahrávky rozpoznat, popřípadě zrychlit rozpoznávací systém tím, že nebudu analyzovat nahrávky, které nevyhovují toleranci příznaku. Filosofie rozpoznávání příznaků je stavěna na tom, že rozpoznávání bylo prováděno v 10 sekundovém, náhodně vyříznutém úseku hudební nahrávky. Příznaky, které byly rozpoznávány, byly tempo nahrávky a dominantní frekvence jednotlivých dob díla.

Cílem tohoto algoritmu, který je popsán na diagramu 2.16 (funkce **beatPerMinuteTracking**) je rozpoznat příznaky nahrávky.

Prvním krokem při analýze nahrávky bylo načtení signálu skladby. Poté byl vyříznut pseudonáhodný úsek 10 sekund skladby.

Nejprve byl určen minimální počáteční čas výřezu, který by měl simulovat realitu. Člověk prakticky chce skladbu rozpoznat až tehdy, zalíbí-li se mu její melodie, což se zřídkakdy stává před desátou sekundou skladby. Minimální počáteční čas výřezu byl tedy nastaven právě na tento čas. Poté byl pomocí generátoru pseudonáhodných

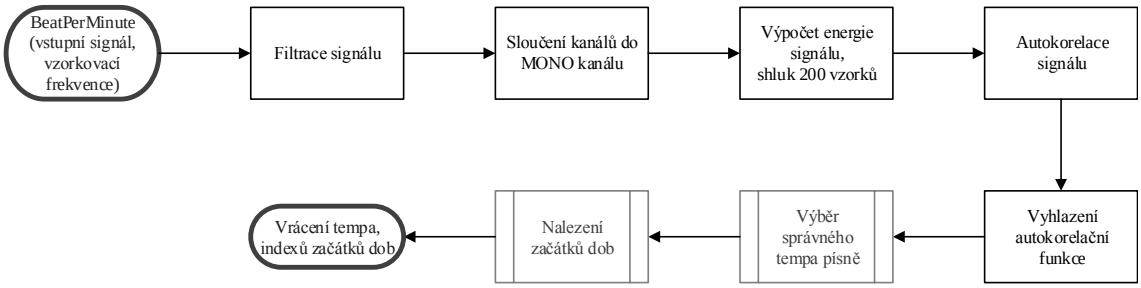


Obrázek 2.16: Algoritmus rozpoznávání příznaků

čísel stanoven vzorek signálu, který stanovil počátek výřezu. Číslo vzorku bylo generováno v rozsahu od desáté sekundy skladby až po desátou sekundu před jejím koncem. Výřez deseti sekund signálu byl vrácen z funkce do hlavního programu. Následně mohlo být přistoupeno k rozpoznání tempa výřezu nahrávky a detekci začátků jednotlivých dob.

2.6.1 Algoritmus BPM

Tento algoritmus slouží k tomu, aby z načtené hudební stopy co nejpřesněji rozpoznal tempo, tedy počet úderů bubeníka za minutu. Proces zachycuje vývojový diagram 2.17 (funkce **BeatPerMinute**).



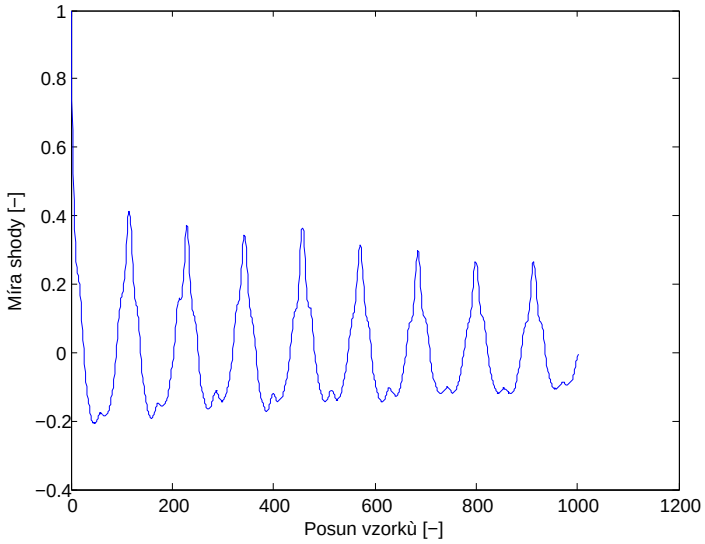
Obrázek 2.17: Postup při zjišťování tempa nahrávky

Prvním krokem byla filtrace signálu pásmovou propustí. Její mezní frekvence byly určeny dříve. Následně bylo provedeno sloučení kanálů signálu do jednoho kanálu. Poté byla vypočtena energie signálu po shlucích 200 vzorků. Shluky nesmí být příliš dlouhé. Je to proto, aby délka shluku nerozmazala energii natolik, že nebude možné detekovat maxima energie. V takovém případě by došlo ke ztrátě klíčové informace z nahrávky.

Dalším krokem byla autokorelace energie. Délku autokorelační funkce jsem zvolil 1000, což při vzorkovací frekvenci 44,1 kHz odpovídá přibližně 4,5 sekundám signálu (berme v úvahu 200 vzorkové shluky). Spodní hranice tempa, vyskytující se v hudebním průmyslu, je 40 úderů za minutu. To je přibližně 0,7 úderu za sekundu.

Znamená to tedy, že pokud by píseň měla opravdu takovéto minimální tempo, bude v grafu autokorelační funkce zaznamenáno 6 vrcholů určujících úder. Z takového počtu lze tempo rozpoznat, proto byla zvolena právě tato délka.

Po provedení autokorelace bylo ještě na funkci aplikováno vyhlazení. To zprůměrovalo zvolených 10 bodů. Vyhladila malá maxima, která by mohla zdánlivě tvořit bicí linku skladby. Příklady autokorelačních funkcí jsou uvedeny na obrázkách 2.18 a 2.19.



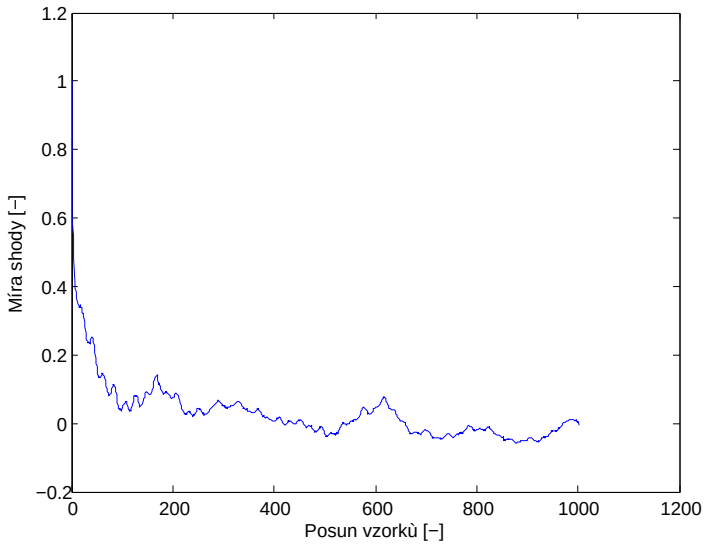
Obrázek 2.18: Příklad autokorelační funkce krátkodobé energie skladby s rozpoznatelným tempem (žánr pop)

Všechny potřebné informace byly připraveny a bylo přistoupeno k samotnému výběru tempa písně, které je popsáno diagramem 2.20 (funkce **chooseCorrectTempo**).

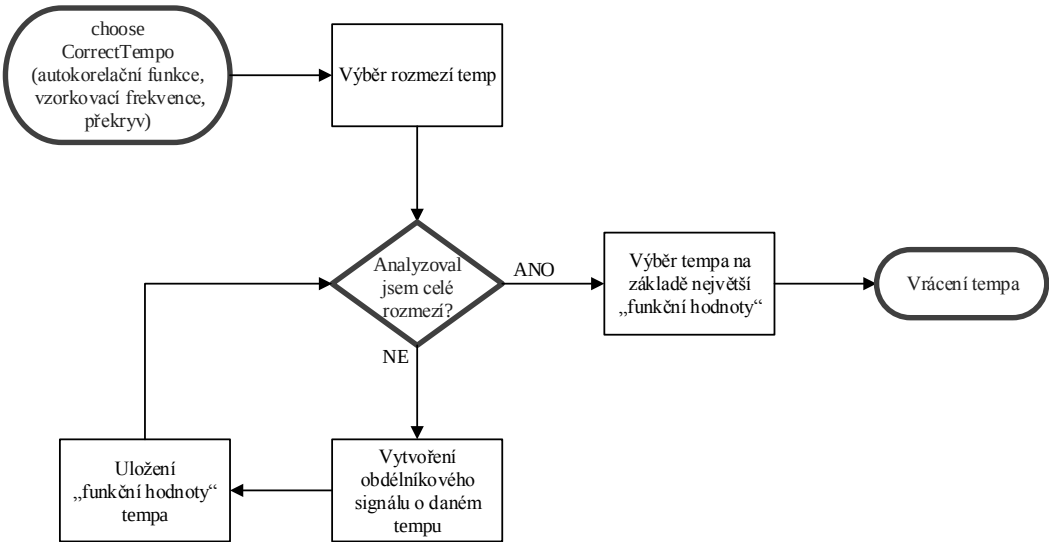
Jak bylo uvedeno výše, tempo bude rozpoznáváno v intervalu 40 až 200 úderů za minutu. Po definici rozmezí byl ve smyčce postupně vytvářen obdélníkový signál (viz obrázek 2.21), který vzestupně simuloval tempo z intervalu. Délka úrovně 1 byla nastavena na 3 vzorky, čímž docílím zpřesnění při rozpoznání. Dalším krokem bylo nalézt funkční hodnotu tempa. Ta je dána vynásobením vektoru autokorelační funkce s vektorem obdélníkového signálu. Pro představu, mají-li oba vektory vrcholy na stejných místech, bude násobená hodnota vyšší, než kdyby se místa neshodovala. Tímto způsobem zjistím funkční hodnoty všech analyzovaných temp. Nalezené tempo je poté definováno jako součin s největší funkční hodnotou.

2.6.2 Detekce jednotlivých dob písně

Po nalezení tempa písně bylo třeba výřez signálu rozdělit podle dob, aby bylo poté možné stanovit příznak každé doby. Tento proces zachycuje vývojový diagram 2.22 (funkce **computeBeatOffset**).



Obrázek 2.19: Příklad autokorelační funkce krátkodobé energie skladby s nerozpoznatelným tempem (žánr klasická hudba)

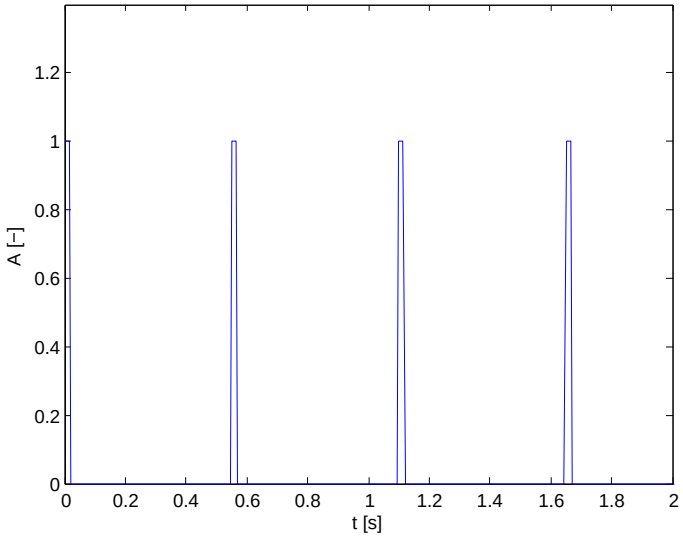


Obrázek 2.20: Vývojový diagram metody tempo pattern

Prvním krokem bylo stanovení délky úrovně 1 obdélníkového signálu, který měl periodu tempa. Délka byla nastavena na hodnotu, která v reálném prostředí odpovídá polovině délky znění úderu bubnu, což je přibližně 0,09 sekundy.

Po vytvoření samotného obdélníkového signálu byla provedena jeho konvoluce s energií signálu (viz obrázek 2.23). Tím byla nalezena místa, kde se vrcholy signálů shodovaly. Bylo-li tempo určeno správně, byly to tedy začátky dob.

Nyní zbývalo tyto začátky upřesnit. Iterativně tedy určujeme první dobu a od



Obrázek 2.21: Generovaný obdélníkový signál

ní hledáme v daném rozmezí, které je dané rozdílem dvou dob, další.

Dalším krokem bylo definovat oblast pro upřesňování indexů počátků dob. Ta byla stanovena na toleranci ±10 BPM, dle nalezeného tempa. Tempo tedy bylo přepočteno na počet vzorků, kde budu začátek doby upřesňovat.

Pak bylo přikročeno k lokalizaci indexů dob. Signál byl tedy postupně procházen směrem k jeho začátku s krokem rozdílu dob. V energii nahrávky bylo v odpovídajícím intervalu nalezeno maximum. To byl hledaný index začátku doby, který byl uložen. Byl nalezen rozdíl mezi dvěma dalšími dobami a postup se opakoval.

Protože jsme se celou dobu pohybovali na úrovni energie, která byla shlukována po 200 vzorkách signálu, bylo třeba po analýze všech dob indexy přepočítat zpět do měřítek signálu. Každý uložený index byl tedy vynásoben 200.

Jelikož jsem indexů začátků lokalizoval zpětně, tedy odzadu, byla celá matice setříděna vzestupně. Výsledná matice indexů byla poté vrácena.

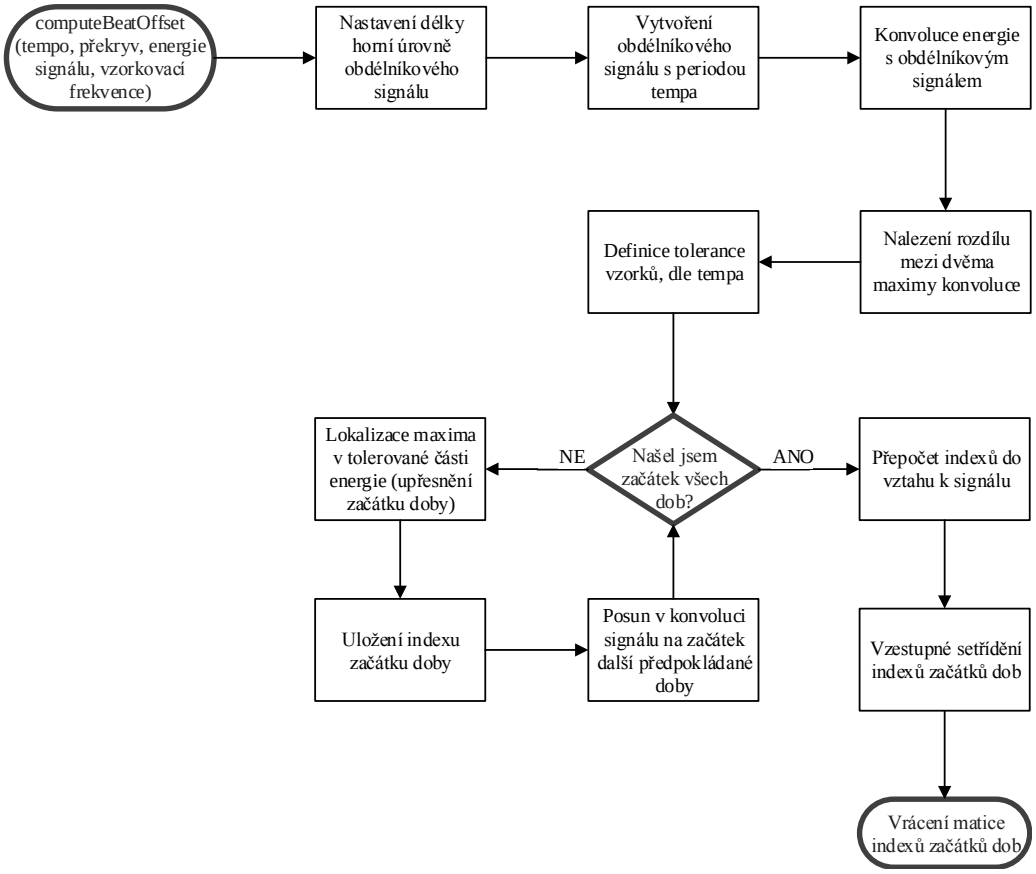
Tím bylo zjištěno tempo písně a začátky dob, se kterými jsem mohl dále pracovat.

2.6.3 Zjištění dominantních frekvencí dob

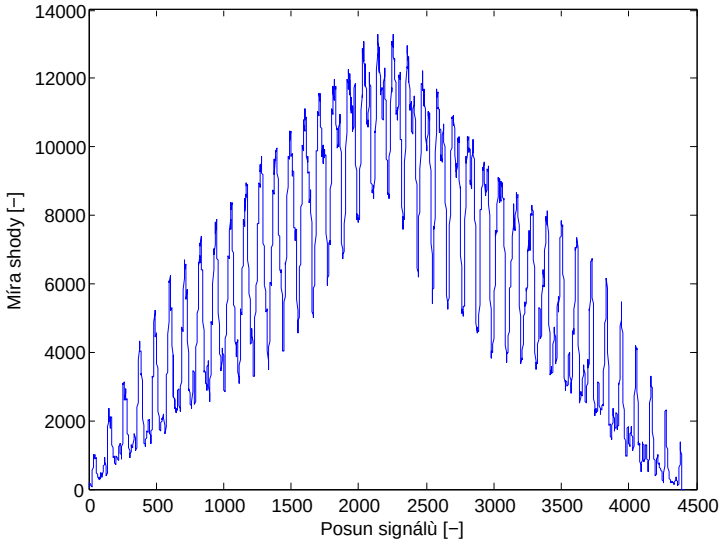
Cílem bylo stanovit určitý počet dominantních frekvencí každé doby, které ji charakterizují. Funkčnost celého systému, která bude popsána v dalších fázích práce, je závislá na počtu frekvencí. Postup při zjišťování dominantních frekvencí dob je zachycen diagramem 2.24 (funkce **getDominantFrequencies**).

Prvním krokem bylo sloučit případný dvoukanálový signál na vstupu do jednoho mono kanálu. Důvody sloučení byly popisovány výše.

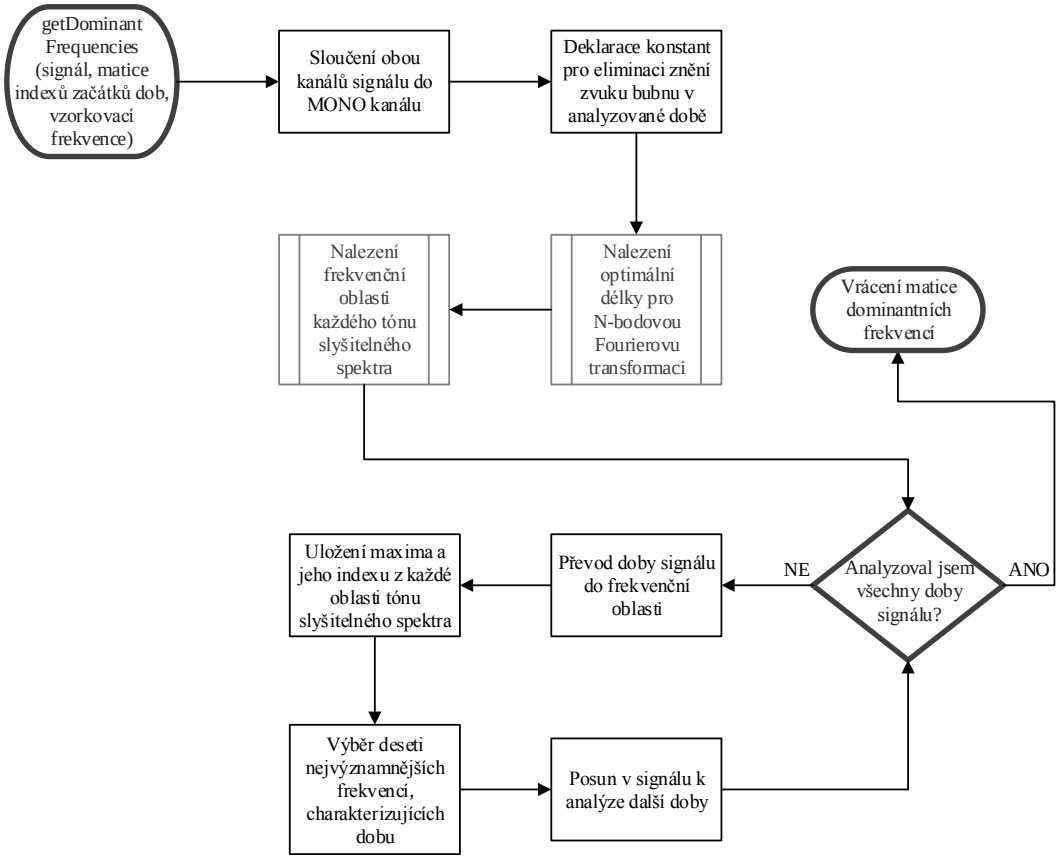
Dále bylo třeba definovat konstanty, které zamezily, aby se do doby dostal zvuk bicích. Tento zvuk obsahuje, podobně jako Diracův impulz, soubor celého spektra



Obrázek 2.22: Postup při detekci dob skladby



Obrázek 2.23: Konvoluce výřezu skladby a obdélníkového signálu



Obrázek 2.24: Určení dominantních frekvencí dob

frekvencí. Kdyby nebyla eliminace provedena, mohla by být za jednu z dominantních frekvencí doby prohlášena právě některá frekvence z bicích, což není správný postup. Doby proto budou zleva ořízntuty o 0,1 sekundy, a zprava o 0,05 sekundy, což naopak eliminovalo chybu bubeníka, tedy brzký úder.

Poté bylo třeba nalézt optimální délku algoritmu FFT, který signál doby převáděl z časové oblasti do frekvenční. Bylo tedy třeba najít dobu, která měla nejdelší počet vzorků. Délka FFT byla potom stanovena jako nejbližší vyšší mocnina dvou.

Dalším krokem byla definice frekvenční osy. Osa byla ořezána pouze na potřebné frekvence, tedy stanovené frekvence, které zařízení reálně přenesou. Následně jsem potřeboval zjistit, jaké indexy na frekvenční ose odpovídaly hranicím tónů. Smyčkou jsem tedy frekvenční osu prošel a stanovil potřebné indexy.

Dále bylo možné přistoupit k analýze dob. Byla načtena doba signálu, která byla ořízntuta o stanovené hranice. Poté byl signál doby převeden do frekvenční oblasti. Třetím krokem bylo vybrání maximální amplitudy z každé oblasti tónu. Bylo-li analyzované rozmezí tónů od G#1-G8, tedy celkový počet 84, byl nalezen stejný počet amplitud. Každá amplituda odpovídala největšímu zesílení z oblasti každého tónu. Spolu s amplitudou byly také ukládány indexy tónů, abych věděl, jaká amplituda přísluší jakému tónu.

Nakonec bylo nutné sestupně setřídit matici amplitud a vybrat určitý počet frek-

vencí, které dobu charakterizují. Jak je uvedeno výše, počet frekvencí byl určen na základě experimentů s množivou vývojových dat. Zjistil jsem, že použití 6 frekvencí je nejvýhodnější. Důvody výběru budou uvedeny v části věnující se samotnému testování systému.

Po zjištění dominantních frekvencí doby bylo přikročeno k další době. Návratová hodnota funkce byla matice obsahující frekvence každé doby.

Tímto byl 10 sekundový výřez hudební nahrávky charakterizován příznaky. Těmi tedy bylo tempo a dominantní frekvence každé doby.

2.7 Testování algoritmu BPM

Pro to, aby bylo zabráněno špatnému vstupu do algoritmů pro detekci dob a dominantních frekvencí, bylo třeba otestovat úspěšnost algoritmu BPM, na jehož výstup oba algoritmy spoléhají. Testováním zjistím procentuální úspěšnost při zjišťování tempa, z kterého mohu vycházet při finálních testech úspěšnosti systému. Kdyby byla finální úspěšnost malá, bylo by zřejmé, že se chyba nacházela právě v algoritmu pro detekci dob, respektive dominantních frekvencí.

Nejprve bylo třeba, abych zjistil tempa všech písní v databázi. Jelikož se v ní nacházejí i písně starší, nebylo možné pomocí dostupných prostředků (především internetu) tempa dohledat. Jelikož se každý interpret a aranž mohou tempem lišit u řady skladeb, přistoupil jsem proto k ručnímu určení tempa. Jediný žánr, který jsem netestoval, byl klasická hudba. Jelikož lze tempa velmi špatně zjistit (nepřítomnost bubeníka), považoval jsem ho za experimentální. Probíhalo to tak, že jsem v programu Audacity do sluchátek pustil část písně, kde byly údery bubeníka zřetelné. Současně jsem spustil nahrávání. Úderem tužky do stolu jsem do mikrofonu notebooku zaznamenal údery bubeníka. Poté jsem změřil průměrnou vzdálenost mezi údery a stanovil tempo.

Po zjištění temp všech písní k databázi jsem ji celou otestoval a výsledky ukládal do aplikace MS Excel. Testování bylo spuštěno nejprve na filtrovaných skladbách, které simulovaly pás frekvencí propustných zařízeními a poté na nefiltrovaných skladbách. Test probíhal obdobně jako v praxi. Pokud bych v reálné aplikaci systému píseň na poprvé nerozpoznal, pokusil bych se ji ještě alespoň dvakrát rozpoznat. Proto test probíhal třikrát. Bylo-li tempo písně ze tří pokusů alespoň jednou rozpoznáno, byl test písně položen za úspěšný. V tabulce 2.3 je znázorněna úspěšnost rozpoznávání BPM.

Výsledky algoritmu u nefiltrovaných skladeb od filtrovaných se ve finále lišily pouze o 0,35%. Z toho lze usoudit, že vliv zařízení na detekci tempa je minimální.

Z tabulky lze vyčíst, že u žánrů, které jsou charakteristické výraznými bicími a stálým rytmem (disco, elektronická hudba, funk, ska), úspěšnost rozpoznání přesahovala 90% i v případě, neberu-li v úvahu dvojité/poloviční tempo.

Faktem je, že u žánru rock, který má poměrně výrazná bicí, bylo rozpoznání 80%. Je to nejspíše proto, že se zde mísí spousta dalších zvučných hudebních nástrojů. Proto bylo s největší pravděpodobností ve zbytku písní tohoto žánru rozpoznáno tempo dvojnásobné nebo poloviční.

Důvodem, proč u žánru hip-hop bylo v 60% případů rozpoznáno dvojité či poloviční tempo je, že bubny jsou prokládány meziúdery, které rozpoznání ztěžují.

Žánry jako country a soul naopak charakterizují nevýrazná bicí, proto má algoritmus nižší úspěšnost rozpoznání přesného tempa. Algoritmus však detekoval dvojnásobné/poloviční tempo.

Je zajímavé, že se rozpoznávání tempa dokázalo vypořádat s triolami v případě žánzu jazz, který měl úspěšnost přes 90%.

Ostatní žánry se pohybovaly kolem 70% až 80% úspěšnosti, v případě rozpoznání polovičních/dvojnásobných temp okolo 10% až 20%.

Další úvahou bylo, jak naložit se skladbami, u kterých algoritmus rozpoznal dvojité či poloviční tempo. Počet skladeb s rozpoznaným dvojitým tempem byl roven 86, poloviční tempo bylo rozpoznáno u 6 skladeb. Z ruční analýzy skladeb bylo potvrzeno, že detekce dvojitého tempa byla oprávněná. Bubeník tedy do bicí linky přidával mezi úder, který byl natolik výrazný, že ho algoritmus detekoval. Rozhodl jsem tedy, že tato chyba vlastně není chybou, protože tempo mezi automaticky parametrizovanou databází a zkoumanou nahrávkou bude detekováno stále shodné. Skladeb s polovičním tempem bylo malé procento. Poslechem bylo usouzeno, že skladby nemají výraznou dobu, tedy, že bicí jsou nevýrazné. Proto je tento problém v podstatě neřešitelný. Algoritmus tedy splnil očekávání, jeho výsledky jsou velmi dobré.

NÁZEV ŽÁNRU	ROZPOZNÁNO [%]	ROZPOZNÁNO DVOJITÉ/POLOVIČNÍ TEMPO [%]	SUMA [%]
blues	76,32	10,53	86,84
country	52,63	42,11	94,74
dechová hudba	71,05	10,53	81,58
disco	100,00	0,00	100,00
elektronická hudba	94,74	0,00	94,74
folk	84,21	7,89	92,11
funk	94,74	2,63	97,37
hip hop	34,21	60,53	89,47
jazz	92,11	5,26	97,37
pop	76,32	21,05	97,37
rnb	81,58	31,16	92,11
reggae	68,42	26,32	92,11
rock	81,58	18,42	100,00
ska	100,00	0,00	100,00
soul	57,89	21,05	76,32
celkem:	**77,72**	**15,96**	**93,68**

Tabulka 2.3: Úspěšnost algoritmu BPM u filtrovaných skladeb

2.8 Uložení matic příznaků do databáze

Další část práce se zabývá uložením příznaků písní do binárních souborů. To bylo třeba provést proto, aby se při každém dotazu na rozpoznání písně nemusely příznaky celé databáze znovu počítat. Dalším kladem tohoto kroku je, že se výrazně zmenší velikost celé databázové struktury wav souborů, a to o 3 řády.

Bylo třeba stanovit strukturu binárního souboru a informace o hudební nahrávce, které bude obsahovat. Struktura je popsána na obrázku 2.25.

FLAG_VECTOR_	DÉLKA NÁZVU ŽÁNRU	DÉLKA NÁZVU PÍSNĚ
char*1	uint16	uint16

FREKVENCE/DOBA	DÉLKA DAT	ŽÁNR	NÁZEV	TEMPO	DATA
uint8	uint32	char*1	char*1	uint8	float32

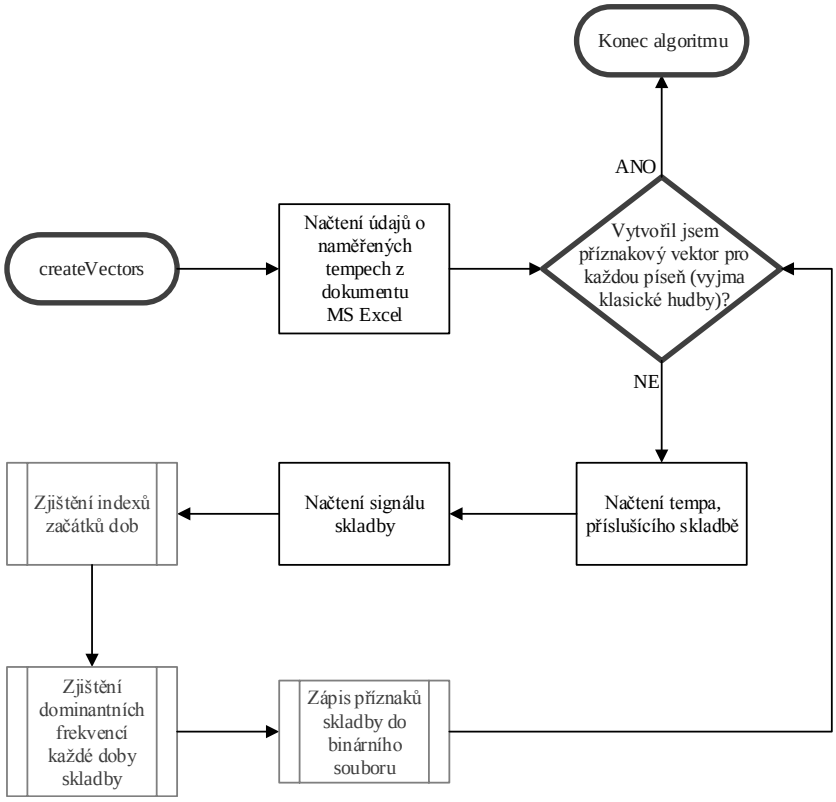
Obrázek 2.25: Struktura binárního souboru

Postup při vytváření databáze binárních souborů zachycuje vývojový diagram 2.26. Vytváření bylo děleno na dvě části. Druhou částí bylo vytváření vektorů klasické hudby. Jelikož tempa nebyla analyzována, tak rozdíl při vytváření binární struktury byl takový, že navíc muselo být zjištěno tempo 10 sekundového výřezu písně, a to pomocí algoritmu BPM (viz vývojový diagram 2.17).

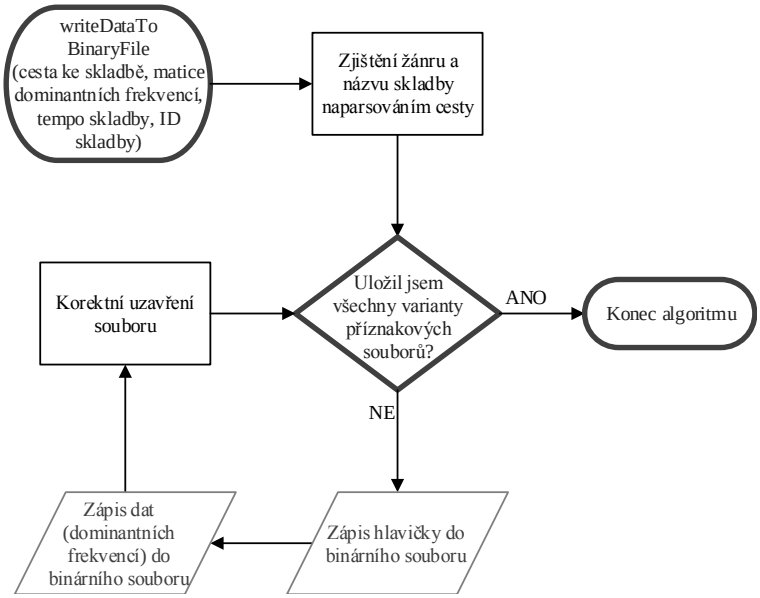
Nejprve bylo třeba načíst naměřené údaje o tempech z dokumentu aplikace MS Excel. Tento krok je možné provést, jelikož tempa uložená v dokumentu již byla jednou rozpoznána, tudíž jsou korektní, a tak měl krok pouze urychlující efekt. Samozřejmě, bylo-li rozpoznáno dvojité či poloviční tempo, bylo použito právě toto tempo. Nebylo-li tempo rozpoznáno vůbec, bylo použito tempo nesprávné. Ve smyčce jsem postupně pro každou píseň (mimo žánru klasické hudby) načítal tempo. Poté byl načten signál skladby. Poté byly zjištěny indexy začátků dob (viz diagram 2.22). Dalším krokem bylo zjištění dominantních frekvencí. Jejich výpočet je popsán vývojovým diagramem 2.24.

Finální částí postupu bylo uložení spočtených příznaků skladby do binárního souboru. Princip je vystižen diagramem 2.27.

Při zápisu dat do souboru byla nejprve naparsována cesta ke skladbě. Tím jsem zjistil jméno interpreta, název písně a žánr skladby. Pro testovací účely, kdy bylo třeba zjistit, jaký počet dominantních frekvencí je při rozpoznávání nejvýhodnější, bylo ukládáno postupně 10 souborů s počtem dominantních frekvencí 1-10. Pro každý soubor byla poté dle obrázku 2.25 zapsána hlavička souboru, poté data s informacemi o dominantních frekvencích. Nakonec byl soubor korektně uzavřen. Po uložení všech souborů pro všechny písně byla databáze příznaků vytvořena.



Obrázek 2.26: Vývojový diagram vytváření binární databáze



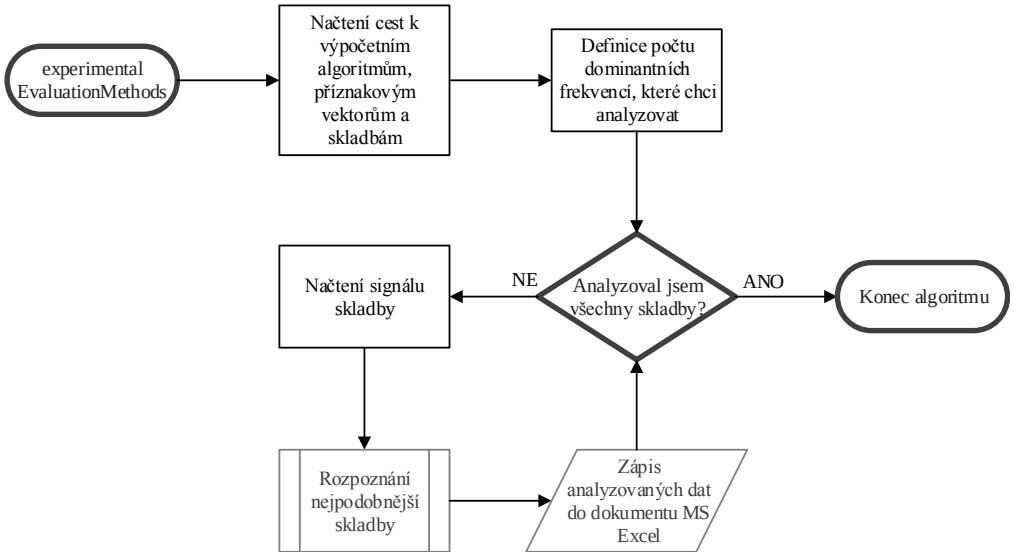
Obrázek 2.27: Zápis příznaků do binárního souboru

2.9 Experimentální vyhodnocovací metody

Poslední části práce bylo vyhodnotit úspěšnost navrženého rozpoznávacího systému. Bylo tedy třeba navrhnout metody, které rozpoznají píseň. Ta je charakte-

rizována příznaky jejího 10 sekundového, pseudonáhodného výřezu. Píseň, kterou obsahuje testovací databáze, je hledána v databázi příznaků, které jsou definovány v binárních souborech.

Celkový průběh vyhodnocení úspěšnosti systému je zobrazen na vývojovém diagramu 2.28.

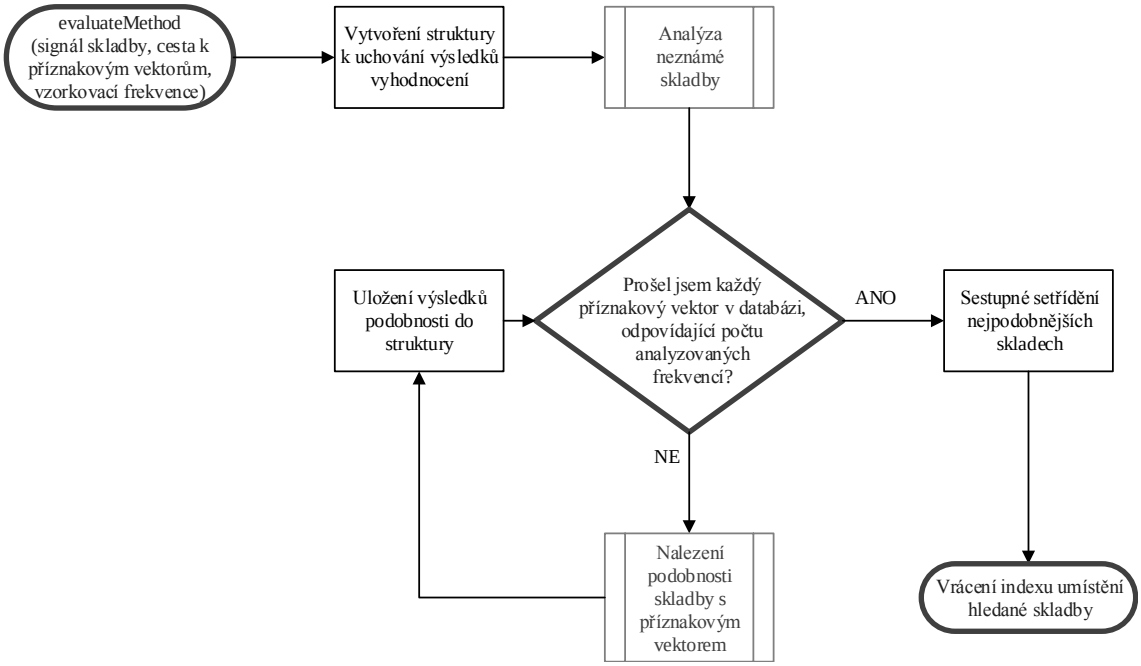


Obrázek 2.28: Postup při experimentálním vyhodnocování úspěšnosti systému

Po načtení cest ke všem potřebným datům bylo třeba definovat počet dominantních frekvencí, které budu používat k analýze. Čím menší počet frekvencí, tím méně informací mám. Na druhou stranu, tím rychleji ale také algoritmy pracují. Bylo tedy třeba nalézt kompromis mezi těmito dvěma faktory. Další skutečností byl fakt, že u některých výřezů písní testovací databáze ani nebyl nalezen takový počet frekvencí, který bych chtěl analyzovat.

Při postupném testování databáze jsem dospěl k závěru, že nejmenší počet detekovaných frekvencí, které skladby obsahovaly, bylo 6. Doba potřebná k rozpoznání skladby se pohybovala kolem 20 sekund na jednom jádře (lze dále paralelizovat), což je přípustná mez. Při testování vzorové databáze bylo 3-7 frekvencí dostačujících k tomu, aby byla skladba rozpoznána a globální úspěšnost neklesala. Při testování reprodukované databáze jsem se tedy chtěl přiklonit k horší variantě, a proto jsem zvolil počet dominantních frekvencí 6.

Poté bylo možné přistoupit k samotné analýze. Postupně jsem tedy načítal varianty skladeb obsažených v testovací (respektive vzorové) databázi (dle varianty testu). Po načtení signálu skladby bylo přikročeno k rozpoznání, které popisuje diagram 2.29. Nejdříve jsem tedy potřeboval vytvořit strukturu, do které bych vhodně uložil výsledky. Ta po vyhodnocení obsahovala údaje o každé skladbě, včetně počtu shodných frekvencí s testovanou nahrávkou. Počet shod charakterizoval skladbu. Nahrávka s největším počtem shod byla označena za nejpodobnější. Poté byla pomocí algoritmů pro výřez 10 sekundového úseku skladby, 2.17, 2.22 a 2.24 analyzována



Obrázek 2.29: Postup při experimentálním vyhodnocování úspěšnosti systému (2)

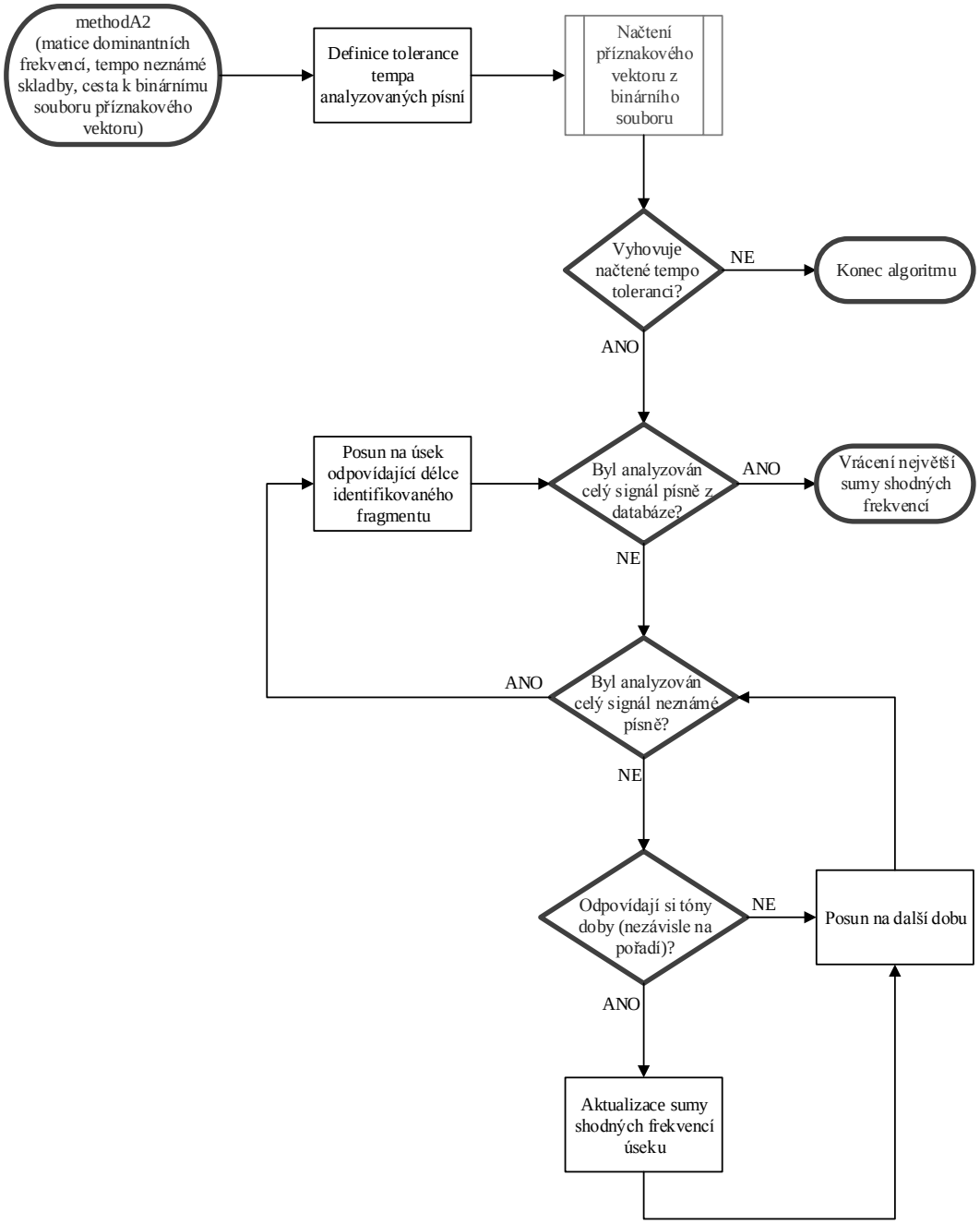
testovaná skladba. Nyní jsem měl k dispozici všechny údaje a mohlo začít postupně naplňování struktury.

Pro testování byly navrženy 2 metody. První metoda pracuje na principu shodných frekvencí, přičemž v rámci doby záleží na amplitudě (pořadí) frekvencí. To znamená, že byla-li frekvence testovaného vzorku ve stejné době na 3. místě a v databázi na 5., nebylo to bráno jako shoda. Jak bylo vyzkoušeno, tato metoda je v reálném prostředí nepoužitelná, a to kvůli vlivu přenosových charakteristik kombinací zařízení.

Proto byla vytvořena druhá metoda, která pracuje obdobně, pouze nezáleží na pořadí dominantních frekvencí v době. Metodu popisují kroky, vyobrazené na diagramu 2.30. Prvním krokem metody byla definice tolerovaného tempa, která byla stanovena na ± 10 BPM. Je to proto, že algoritmus pro nalezení tempa má schopnost vypořádat se právě s touto odchylkou tempa.

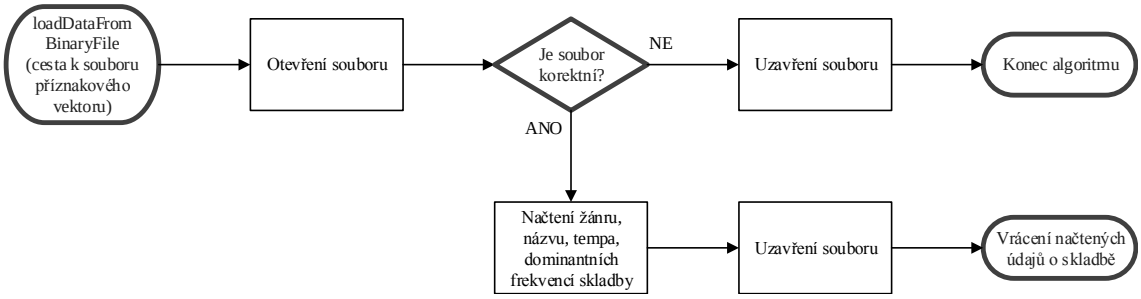
Poté bylo třeba načíst příznakový vektor, který charakterizoval píseň uloženou v databázi. Tato funkce je popsána diagramem 2.31. Následně bylo třeba zjistit, zda rozdíl analyzovaných temp vyhovuje toleranci. Pokud ne, byl algoritmus ukončen a bylo přistoupeno k vyhodnocení další písně. Pokud tempo vyhovovalo, začalo porovnávání. Příznaky písně z databáze byly postupně porovnávány s příznaky 10 sekundového výřezu testované nahrávky. Pokud si tóny v době odpovídaly, byla inkrementována konečná suma společných frekvencí v daném posunu a přistoupilo se k analýze další doby. Takto se postupovalo, dokud jsem neporovnal testovaný výřez s celou písní. Posledním krokem funkce byl výběr největší sumy shodných frekvencí.

Suma, žánr a název písně byly uloženy do struktury. Po analýze celé databáze byla naplněná struktura sestupně setříděna, dle počtu společných frekvencí. Nako-



Obrázek 2.30: Vývojový diagram porovnávací funkce

nec bylo vráceno číslo, charakterizující, kolikátý nejpodobnější byl vektor testované písně. Následoval zápis výsledků analýzy do souboru.



Obrázek 2.31: Načítání příznaků z binárního souboru

3 Dílčí vyhodnocení a postřehy

Celkem byly provedeny tři druhy dílčích vyhodnocení funkčnosti systému. Mohl jsem tedy zjistit, jak si systém poradil s různě modifikovaným nahrávacím řetězcem a určit, jaký vliv má modifikace na funkčnost systému. Modifikací řetězce je rozuměno zahrnutí reálných nahrávacích podmínek.

Samotné testování je časově náročné. Porovnání jednoho 10 sekundového výřezu nahrávky s celou databází, včetně zápisu výsledků do aplikace MS Excel, je průměrně 17 sekund. Jedním z důvodů je právě zápis výsledků. Integrovaná funkce v programu Matlab, která zápis řeší, si bere příliš systémových prostředků a zpomaluje celý proces.

Mimo toho se stávalo, že celý testovací proces byl po určité době, která se pohybovala přibližně v intervalu 40 písní, systémem zastaven, a to právě z důvodu nedostatku prostředků pro zápis výsledků. Byl jsem tedy nucen testování pustit znovu od poslední testované písně.

Dále je třeba podotknout, že každý test, také z časových důvodů, probíhal pouze jednou. Nastala-li tedy situace, kdy tempo písně nebylo ve vyříznutém úseku rozpoznáno správně, i když v jiném úseku by rozpoznáno správně bylo (tento fakt vyplynul z testu algoritmu BPM), byla správná rozpoznávací schopnost systému značně omezena. Tím bych chtěl tedy podotknout, že při více opakovaných testech rozpoznání písní by výsledky systému byly vyšší a objektivnější. Vycházím také z analogie reality, kdy při nenalezení písně pokus opakujeme vícekrát. Tok času nám poté zabrání tomu, abychom pro nahrávání vybrali stejný časový úsek písně.

Technické vybavení počítače pro testování systému bylo následující: procesor Intel Core i5, 2.30 GHz, 4 GB RAM, integrovaná grafická karta Intel HD Graphics 3000. Na testovacím počítači byl nainstalován operační Windows 8.1 Pro N, 64 bit.

3.1 Testování výřezu originální nahrávky

Prvním testem, který by měl mít největší úspěšnost při rozpoznání, je testování originální nahrávky vůči originální nahrávce. To znamená, že vliv nahrávacího prostředí nebyl do nahrávky zahrnut. Nahrávky byly tedy testovány tak, jak byly do databáze ukládány. Výsledky testování vystihuje tabulka 3.1.

Dále je třeba dodat, že v tomto testu byla testována celá databáze vůči celé databázi příznakových vektorů.

NÁZEV ŽÁNRU	Počet analyzovaných frekvencí [-]																x̄ (žánry) [%]
	10				*5*				*3*				*1*				
	Rozpoznáno v TOP x [%]																
	TOP 1	TOP 3	TOP 5	TOP 10	TOP 1	TOP 3	TOP 5	TOP 10	TOP 1	TOP 3	TOP 5	TOP 10	TOP 1	TOP 3	TOP 5	TOP 10	
blues	84,2%	86,8%	86,8%	86,8%	86,8%	89,5%	89,5%	89,5%	81,6%	86,8%	89,5%	89,5%	78,9%	86,8%	86,8%	89,5%	86,8%
country	92,1%	94,7%	94,7%	97,4%	84,2%	89,5%	89,5%	92,1%	89,5%	92,1%	94,7%	94,7%	94,7%	94,7%	94,7%	97,4%	92,9%
dechová hudba	78,9%	81,6%	84,2%	92,1%	76,3%	78,9%	78,9%	81,6%	81,6%	84,2%	84,2%	84,2%	68,4%	73,7%	81,6%	86,8%	81,1%
disco	89,5%	92,1%	94,7%	94,7%	86,8%	94,7%	97,4%	97,4%	92,1%	94,7%	94,7%	97,4%	78,9%	89,5%	97,4%	100,0%	93,3%
elektronická hudba	76,3%	92,1%	94,7%	97,4%	81,6%	94,7%	94,7%	94,7%	73,7%	81,6%	89,5%	97,4%	42,1%	57,9%	63,2%	68,4%	81,3%
folk	81,6%	84,2%	86,8%	94,7%	86,8%	92,1%	94,7%	97,4%	78,9%	92,1%	92,1%	94,7%	86,8%	89,5%	94,7%	97,4%	90,3%
funk	84,2%	89,5%	92,1%	94,7%	92,1%	92,1%	92,1%	94,7%	76,3%	89,5%	89,5%	94,7%	78,9%	92,1%	97,4%	97,4%	90,5%
hip hop	81,6%	86,8%	89,5%	94,7%	94,7%	94,7%	97,4%	97,4%	89,5%	94,7%	94,7%	97,4%	94,7%	94,7%	100,0%	100,0%	93,9%
jazz	89,5%	92,1%	92,1%	94,7%	94,7%	94,7%	94,7%	97,4%	84,2%	92,1%	92,1%	97,4%	89,5%	92,1%	92,1%	100,0%	93,1%
klasická hudba	47,4%	47,4%	47,4%	47,4%	31,6%	31,6%	31,6%	31,6%	34,2%	34,2%	34,2%	34,2%	23,7%	26,3%	26,3%	26,3%	34,7%
pop	94,7%	97,4%	97,4%	97,4%	92,1%	97,4%	97,4%	100,0%	89,5%	97,4%	97,4%	97,4%	94,7%	94,7%	94,7%	94,7%	95,9%
rnb	89,5%	92,1%	92,1%	92,1%	89,5%	89,5%	97,4%	97,4%	92,1%	92,1%	92,1%	92,1%	89,5%	92,1%	92,1%	94,7%	92,3%
reggae	89,5%	92,1%	92,1%	94,7%	92,1%	94,7%	94,7%	97,4%	86,8%	89,5%	92,1%	94,7%	94,7%	94,7%	94,7%	94,7%	93,1%
rock	94,7%	94,7%	94,7%	94,7%	89,5%	89,5%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	84,2%	97,4%	97,4%	97,4%	95,9%
ska	97,4%	97,4%	97,4%	100,0%	97,4%	97,4%	100,0%	100,0%	84,2%	89,5%	89,5%	97,4%	89,5%	94,7%	97,4%	97,4%	95,4%
soul	60,5%	76,3%	78,9%	86,8%	73,7%	73,7%	76,3%	86,8%	73,7%	78,9%	78,9%	84,2%	63,2%	73,7%	81,6%	84,2%	77,0%
x̄ (TOP x) [%]	83,2%	87,3%	88,5%	91,3%	84,4%	87,2%	89,1%	91,0%	81,7%	86,8%	87,8%	90,5%	78,3%	84,0%	87,0%	89,1%	
x̄ (frekvence) [%]	87,6%				87,9%				86,7%				84,6%				

Obrázek 3.1: Výsledky testování výřezu originální nahrávky vůči databázi příznakových vektorů

3.1.1 Vyhodnocení testu výřezu originální nahrávky

Z tabulky lze vyvodit zajímavé závěry. Z pohledu žánrů má nejhorší úspěšnost klasická hudba. Důvod vychází ze špatné analýzy tempa. To je ostatně vidět i v tabulce 2.3, kde je znázorněna úspěšnost nalezení správného tempa písně.

Další žánr, který se od ostatních výrazněji liší nízkou úspěšností, je soul, který má průměrně 77% úspěšnost nalezení skladby. Ten je také charakterizován ne příliš výraznou bicí linkou.

Dalšími žánry, které se průměrně nedostaly přes 90% nalezených skladeb, jsou blues, country a elektronická hudba. U elektronické hudby, která se naopak vyznačuje výraznými bicími, hrál špatný faktor počet analyzovaných frekvencí. Ostatní žánry byly rozpoznány s úspěšností větší, než 90%.

Z pohledu počtu analyzovaných frekvencí se ukázalo, že všeobecně nejlepší výsledky jsou s pěti dominantními frekvencemi za dobu. S tímto počtem frekvencí se napříč žánry dostalo pro TOP 1 84,4% úspěšnosti, což je uspokojivé číslo.

I tento fakt byl jeden z důvodů, proč jsem se rozhodl při analýze vlivů zařízení a impulzních odezev místností spoléhat na 6 dominantních frekvencí za dobu.

3.2 Testování výřezu reprodukované nahrávky ve výchozím prostředí

Druhým testem bylo zahrnout do nahrávek přenosové vlastnosti testovaných kombinací zařízení. Tím jsem zjistil, jaký vliv mají zařízení na rozpoznávací systém. Tabulka s výsledky testu je na obrázku 3.2 a 3.3. Otestována byla opět celá databáze.

			Reprodukční zařízení			
			Notebook	*Projektor*	*2.1 reproduktory*	
			Asus K50ID	*BenQ MS-500H*	*Creative*	*Creative*
			-	-	*se subwooferem*	*bez subwooferu*
Nahr. z.	*Mobilní telefon*	*Apple iPhone 4*	8,55%	6,58%	12,50%	47,37%
	Náhlavní sluchátka	*GemBird HeadPhones*	14,14%	12,50%	23,03%	42,11%
	Notebook	*HP Probook 4530s*	12,17%	23,03%	46,88%	35,36%

Obrázek 3.2: Výsledky testování reprodukované nahrávky ve výchozím prostředí (1)

			Reprodukční zařízení		
			2.1 reproduktory		*Mobilní telefon*
			GX Gaming SW-G2.1 1250	*GX Gaming SW-G2.1 1250*	*Samsung GSH-i900*
			se subwooferem	*bez subwooferu*	-
Nahr. z.	*Mobilní telefon*	*Apple iPhone 4*	46,55%	46,22%	27,14%
	Náhlavní sluchátka	*GemBird HeadPhones*	45,23%	42,43%	22,37%
	Notebook	*HP Probook 4530s*	51,32%	19,08%	6,58%

Obrázek 3.3: Výsledky testování reprodukované nahrávky ve výchozím prostředí (2)

Je nutné zmínit, že veškeré výsledky jsou TOP 10. To znamená, že pokud se nahrávka vešla do 10 nepodobnějších, považoval jsem nalezení za úspěšné. Je to tak proto, že skóre nalezení by bylo příliš nízké. Z tabulek by poté nešly vyvodit závěry.

Z výsledků vyobrazených v tabulce je vidět, že už samotné přenosové charakteristiky testovaných zařízení mají na nahrávací řetězec a následnou funkčnost systému vliv. Nejvyššího rozpoznávacího skóre dosáhla kombinace zařízení 2.1 reproduktorů značky GX Gaming, se subwooferem a mikrofon notebooku.

Z pohledu reprodukčních zařízení měly největší průměrnou rozpoznávací schopnost 2.1 reproduktory GX Gaming, v kombinaci se zapnutým subwooferem. Celkově lze tvrdit, že reproduktory byly z testovaných zařízení nejspolehlivější.

Všechna tři nahrávací zařízení měla průměrnou úspěšnost kolem 31%, a tak se v tomto testu zdají být stejně schopná.

3.3 Testování výřezu reprodukované nahrávky s impulzní odezvou místnosti

Posledním testem bylo zjistit, jakou úspěšnost bude mít rozpoznávací systém na reprodukované nahrávky, spolu s impulzní odezvou místnosti. Zde nebyl z důvodů časové náročnosti proveden kompletní test celé databáze, ale 28 písní na žánr, čili celkem 448 písní. Z důvodu, aby výsledky vyhodnocení nemohl druh žánru ovlivnit více, než jiný, byl testován stejný počet písní pro každý žánr.

Vyhodnocení testu pro místnost školní učebny v Londýně je znázorněno na obrázcích 3.4 a 3.5. Výsledky Opera House v Sydney jsou na obrázcích 3.6 a 3.7. Shrnutí testu pro koupelnu v Schulz Building ve městě Adelaide je ilustrováno na obrázcích 3.8 a 3.9.

3.3.1 Výsledky systému pro impulzní odezvu školní učebny v Londýně

Z tabulky lze vyčíst, že žádná kombinace zařízení nedokázala rozpoznat více, než 31,19% písní. Tomu tak bylo v případě kombinace 2.1 reproduktorů a mobilního telefonu. Nejhoršího výsledku dosáhla kombinace reproduktoru projektoru a mobilního telefonu.

Z pohledu reprodukčních zařízení nejlépe dopadly obecně obě dvě testované sestavy 2.1 reproduktorů (vyjma GX Gaming, bez subwooferu), které průměrně překonaly 21% úspěšnost rozpoznání. Ostatní se pohybovaly kolem 13% procentům úspěšnosti rozpoznání.

Nejlepšími nahrávacími zařízeními byly náhlavní sluchátka a mikrofon mobilního telefonu. Průměrná schopnost rozpoznání byla 15%.

Celkově tedy lze zhodnotit, že vliv této místnosti na nahrávací řetězec je velký.

Výsledky těchto testů jsou opět TOP 10.

3.3.2 Výsledky systému pro impulzní odezvu Opera House v Sydney

U této impulzní odezvy dopadly testy obdobným způsobem. Individuálně největší skóre měly 2.1 reproduktory, opět vyjma sestavy GX Gaming, bez subwooferu. Úspěšnost se pohybovala kolem 33%.

Nejlepšími reprodukčními zařízeními byly GX Gaming reproduktory, se subwooferem a Creative reproduktory, také v kombinaci se subwooferem. Zařízením s nejhorší rozpoznávací schopností byl reproduktor mobilního telefonu.

Nahrávacími zařízeními s největší úspěšností rozpoznání byly sluchátka a notebook. Průměrné skóre je 18%.

Tato místnost také nahrávky ovlivní natolik, že systém nemůže písně rozpoznat s vyšší úspěšností. Z tabulky je vidět, že ani tak nezáleženo na schopnostech nahrávacích zařízení, jako na přenosových charakteristikách zařízení reprodukčních.

3.3.3 Výsledky systému pro impulzní odezvu koupelny v Schulz Building

Tyto testovací parametry zaznamenaly největší úspěšnost při rozpoznání. Nejvyšší zaznamenané skóre v individuální kombinaci měly 2.1 reproduktory Creative a náhlavní sluchátka, a to necelých 50%.

Tabulka opět popisuje, že nejlepšími nahrávacími zařízeními, stejně jako u ostatních případů impuzních odezev, byly 2.1 reproduktory (vyjma GX Gaming, bez subwooferu). Jejich průměrná úspěšnost se pohybovala kolem 40%.

Opět se ukazuje, že majoritní jsou schopnosti reprodukčních zařízení. Z nahrávacích zařízení se nejlépe jeví náhlavní sluchátka a notebook, průměrně mají 21% úspěšnost rozpoznání.

Impulzní odezva místnosti znovu velmi ovlivňuje schopnost systému rozpoznávat písně, ikdyž ze všech tří testovaných místností nejméně.

3.3.4 Vyhodnocení testu výřezu reprodukované nahrávky s impulzní odezvou místnosti

K testu lze obecně říci, že impulzní odezva místnosti velmi ovlivní nahrávací řetězec. Z tabulek lze také vyvodit závěr, že záleží spíše na schopnosti reprodukčních zařízení.

Školní učebna univerzity v Londýně			Reprodukční zařízení			
			Notebook	Projektor	2.1 reproduktory	
			Asus K50ID	BenQ MS-500H	Creative	Creative
			-	-	se subwooferem	bez subwooferu
Nahr. z.	Mobilní telefon	Apple iPhone 4	11,06%	10,62%	24,34%	19,47%
	Náhlavní sluchátka	GemBird HeadPhones	12,39%	13,94%	19,03%	25,22%
	Notebook	HP Probook 4530s	13,27%	13,94%	20,35%	21,90%

Obrázek 3.4: Výsledky testování výřezu reprodukované nahrávky pro školní učebnu v Londýně (1)

Školní učebna univerzity v Londýně			Reprodukční zařízení		
			2.1 reproduktory		Mobilní telefon
			GX Gaming SW-G2.1 1250	GX Gaming SW-G2.1 1250	Samsung GSH-i900
			se subwooferem	bez subwooferu	-
Nahr. z.	Mobilní telefon	Apple iPhone 4	31,19%	13,72%	12,17%
	Náhlavní sluchátka	GemBird HeadPhones	19,03%	18,58%	12,17%
	Notebook	HP Probook 4530s	16,81%	13,50%	12,39%

Obrázek 3.5: Výsledky testování výřezu reprodukované nahrávky pro školní učebnu v Londýně (2)

Opera House v Sydney			Reprodukční zařízení			
			Notebook	Projektor	2.1 reproduktory	
			Asus K50ID	BenQ MS-500H	Creative	Creative
			-	-	se subwooferem	bez subwooferu
Nahr. z.	Mobilní telefon	Apple iPhone 4	9,29%	10,18%	34,07%	19,91%
	Náhlavní sluchátka	GemBird HeadPhones	10,40%	13,05%	33,63%	35,40%
	Notebook	HP Probook 4530s	12,17%	13,50%	30,97%	33,41%

Obrázek 3.6: Výsledky testování výřezu reprodukované nahrávky pro Opera House v Sydney (1)

Opera House v Sydney			Reprodukční zařízení		
			2.1 reproduktory		Mobilní telefon
			GX Gaming SW-G2.1 1250	GX Gaming SW-G2.1 1250	Samsung GSH-i900
			se subwooferem	bez subwooferu	-
Nahr. z.	Mobilní telefon	Apple iPhone 4	34,07%	13,72%	7,96%
	Náhlavní sluchátka	GemBird HeadPhones	34,51%	15,93%	8,41%
	Notebook	HP Probook 4530s	31,19%	15,04%	10,40%

Obrázek 3.7: Výsledky testování výřezu reprodukované nahrávky pro Opera House v Sydney (2)

Schulz Building ve městě Adelaide (Austrálie)			Reprodukční zařízení			
			Notebook	Projektor	2.1 reproduktory	
			Asus K50ID	BenQ MS-500H	Creative	Creative
			-	-	se subwooferem	bez subwooferu
Nahr. z.	Mobilní telefon	Apple iPhone 4	8,19%	7,30%	46,02%	29,42%
	Náhlavní sluchátka	GemBird HeadPhones	9,29%	9,96%	39,16%	49,56%
	Notebook	HP Probook 4530s	9,51%	15,49%	37,83%	42,48%

Obrázek 3.8: Výsledky testování výřezu reprodukované nahrávky pro koupelnu v Schulz Building (1)

Schulz Building ve městě Adelaide (Austrálie)			Reprodukční zařízení		
			2.1 reproduktory		Mobilní telefon
			GX Gaming SW-G2.1 1250	GX Gaming SW-G2.1 1250	Samsung GSH-i900
			se subwooferem	bez subwooferu	-
Nahr. z.	Mobilní telefon	Apple iPhone 4	46,24%	13,27%	7,30%
	Náhlavní sluchátka	GemBird HeadPhones	36,06%	22,57%	7,96%
	Notebook	HP Probook 4530s	37,61%	15,49%	10,18%

Obrázek 3.9: Výsledky testování výřezu reprodukované nahrávky pro koupelnu v Schulz Building (2)

4 Závěr

Databáze nahrávek pro vývoj a testování systému byla vytvořena. Její velikost bylo 608 nahrávek, rozdělených do 16 nejběžnějších světových žánrů. Na jeden žánr tedy připadlo 38 hudebních nahrávek.

Byl úspěšně navržen postup pro určení a nahrávání úryvků přehrávaných a zaznamenaných několika různými zařízeními v různém prostředí. Celkem bylo testováno 7 reprodukčních a 3 nahrávací zařízení, což dohromady vytvořilo 21 různých testovacích kombinací zařízení. Byly naměřeny jejich přenosové charakteristiky a vytvořeny filtry každého zařízení. Různé prostředí bylo vytvořeno pomocí volně dostupných impulzních odezev 3 místností s různými akustickými vlastnostmi.

Vlastní simulace proběhla pomocí vytvoření nové, testovací databáze, která obsahovala všechny kombinace písní se zařízeními (zakomponování filtrů do nahrávek) a impulzních odezev prostředí. Byla tedy vytvořena databáze o velikosti 38304 hubeních nahrávek.

Byl vytvořen vlastní systém pro identifikaci hudební nahrávky z databáze podle jejího úryvku. Délka úryvku byla rovna 10 sekundám náhodného výřezu skladby. Systém pracoval na principu beat synchronous. Naprogramovanými metodami bylo zjišťováno tempo nahrávky. Výřez hudební nahrávky byl následně rozdělen na jednotlivé doby a byly určeny dominantní frekvence dob. Pro zajištění správného postupu bylo vše podloženo testy.

Výsledky testu algoritmu BPM byly uspokojivé. Dosažené skóre bylo 78% při toleranci ± 10 BPM. Tu však algoritmy umějí vyrovnat, a tak lze tempo v toleranci považovat za správný výsledek. Spolu s rozpoznaným poloviční/dvojitým tempem byla úspěšnost algoritmu 94%, a to při stejné toleranci.

Dvojitá či poloviční tempa nebylo ve finále nutné považovat za chybu. Detekce tohoto tempa byla po sluchové analýze oprávněná. Bubeník buď přidal do rytmu mezi úder, nebo zahrál dobu nevýrazně.

Tempo každé písně z hudební databáze bylo nejprve analyzováno ručně, protože záznamy o tempech starších písní nebyly na internetu dostupné.

Pro rychlejší práci systému a redukci velikosti databáze byly vypočtené příznakové vektory uloženy do databáze binárních souborů. Ty byly při testování postupně načítány a porovnávány s hledaným úryvkem písně.

Pomocí jedné (časové důvody) z navržených testovacích metod byly provedeny tři druhy testů.

Prvním byl test databáze originální databáze. Průměrná úspěšnost rozpoznání žánrů se pohybovala kolem 90%. Nejnižší úspěšnost měl žánr klasické hudby, u kterého se podařilo rozpoznat pouze 35% nahrávek. Navržený systém si tedy na origi-

nálních nahrávkách vedl dobře. Tento test byl důležitý také z toho důvodu, že z něj stanoveny parametry pro další testování.

Druhým testem bylo zahrnutí přenosových vlastností do nahrávek. Zde byl faktor přenosových vlastností testovaných zařízení znatelný. Nejvyšší rozpoznávací úspěšnost byla 52%.

Třetím testem bylo ještě navíc zahrnout do nahrávek impulzní odezvy testovaných místností. Bylo znát, že tento faktor má na nahrávací řetězec velký vliv. Zde se úspěšnosti nalezení pochybovaly na hranici 35%, a to pouze u úspěšnějších kombinací testovaných zařízení. Byl také nalezen fakt, že systém schopen rozpoznat spíše skladby, které mají bicí linku na dobře slyšitelné úrovni (tím se také beat synchronous systémy vyznačují).

Do budoucna by bylo dobré zlepšit rozpoznávací systém tak, aby lépe detekoval tempo za ztížených podmínek (přenosové vlastnosti zařízení, impulzní odezva místnosti), a nedocházelo tak k nerozpoznání tempa. Dalším krokem by bylo naprogramovat práh, který by určil, kdy byla píseň rozpoznána a kdy ne. Systém by tedy nevracel číslo, popisující pořadí podobnosti písně, ale název písně nebo informaci o tom, že píseň rozpoznána nebyla.

Literatura

[1] CASEY, M. A. *Content-Based Music Information Retrieval: Current Directions and Future Challenges*. Proceedings of the IEEE, 2008. ISBN 0018-9219.

[2] HAYES, M. H. *Schaum's Outlines of Digital Signal Processing*. McGraw-Hill international editions. Statistics series. McGraw-Hill, 1999. ISBN 0-07-027389-8.

[3] HAYES, M. H. *Statistical digital signal processsing and modeling*. John Wiley & Sons, Inc., 1996. ISBN 0-471-59431-8.

[4] PORAT, B. *A Course in Digital Signal Processing*. John Wiley & Sons, Inc., 1997. ISBN 0-471-14961-6.

[5] WIKIBOOKS.ORG. *Praktická elektronika/Spektrum signálu* [online]. 2011. Dostupné z: `<cs.wikibooks.org/wiki/Praktická_elektronika/Spektrum_signálu>`.

[6] WIKIPEDIA.ORG. *Kategorie: Zpracování signálu* [online]. 2014. Dostupné z: `<http://cs.wikipedia.org/wiki/Kategorie:Zpracování_signálu>`.

[7] ZAPLATíLEK KAREL, D. B. *Matlab - začínáme se signály*. Ben, 2010. ISBN 978-80-7300-200-0.

A Obsah přiloženého DVD

- Veškeré zdrojové kódy této práce,
- dokumenty se zdrojovými daty testů a jejich výsledky,
- PDF podoba tohoto dokumentu,
- textový seznam všech testovaných písní,
- graficky zpracované přenosové charakteristiky testovaných zařízení,
- zvukové soubory impulzních odezev testovaných místností,
- zvukové soubory přenosových charakteristik testovaných zařízení,
- generované signály slyšitelného zvukového spektra,
- soubory aplikace LaTeX, použité při vytváření tohoto dokumentu.